Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
HUMAN RESOURCES RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *Human Resources Research Center*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

THE JOHN CRERAR
OCT 21 1952
LIBRARY

CONTENTS

The Description of Facial Expressions in Terms of Two Dimensions: H. SCHLOSBERG 229

Remote Action Potentials at the Moment of Response in a Simple Reaction-Time Situation: R. L. HENDERSON 238

Different Strengths of Set in the Solution of Anagrams: I. MALTZMAN AND L. MORRISETT, JR. 242

Effect of Varying Amounts of Rest on Conventional and Bilateral Transfer "Reminiscence": G. R. GRICE AND B. REYNOLDS 247

On the Limits of Stereoscopic Vision: K. N. OGLE 253

Information Theory and Immediate Recall: M. ABORN AND H. RUBENSTEIN 260

Instructional Sets and Habit Interference: M. H. SMITH, JR. 267

Nonindependence of Successive Responses in Measurements of the Visual Threshold: W. S. VERPLANCK, G. H. COLLIER, AND J. W. COTTON 273

Factors Related to Speed and Accuracy of Tactual Discrimination: T. R. AUSTIN AND R. B. SLEIGHT 283

Functional Fixedness as Related to Problem Solving: A Repetition of Three Experiments: R. E. ADAMSON 288

Simultaneous Vision and Audition: The Detection of Elements Missing from Overlearned Sequences: G. H. MOWBRAY 292

American Psychological Association

Vol. 44 No. 4 October 1952

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.00, or $14.00 annually. Single copies are $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1515 Massachusetts Ave. N.W., Washington 5, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors who request them are supplied with 50 free reprints without covers. All of the cost of an author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas. Priority in publication is given to articles whose authors assume the full cost of publication.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 (4, Suppl.), 389–449. In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Journal of
Experimental Psychology

VOL. 44, NO. 4 OCTOBER, 1952

THE DESCRIPTION OF FACIAL EXPRESSIONS IN TERMS OF TWO DIMENSIONS

HAROLD SCHLOSBERG

Brown University

Early studies on facial expression stressed the inaccuracy of judgments. In 1938 Woodworth suggested that the inaccuracy was more apparent than real, in that it resulted from scoring judgments on a simple *right-wrong* basis. Through an analysis of published results from earlier studies, he developed a scale, consisting of the following six steps: (1) Love, Happiness, Mirth; (2) Surprise; (3) Fear, Suffering; (4) Anger, Determination; (5) Disgust; (6) Contempt. There was also a seventh category, Scattering, for pictures which didn't fit into the other six. The use of such a scale showed clearly that *S*s made many minor errors, but few major ones. That is, they would frequently transpose a picture from Step 3 to Step 4, but they were less apt to misplace it in more remote steps, as 1 or 6.

Schlosberg (**6**) tested this scale on a new series of pictures, those of Frois-Wittmann (**1**), published by Hulin and Katz (**3**). The pictures were sorted into a linear series of bins, labeled in terms of the Woodworth categories. It was immediately obvious that pictures whose modes fell in Step 6 were as apt to spread over into Step 1 as into Step 5. This could only mean that the scale was recurrent, rather than linear. Hence, the computations of mean scale positions and average deviations were based on the assumption that the scale was circular; that is, a picture which was evenly split into Steps 6 and 1 was assigned a mean of 6½, rather than of 3½, which would have put it on the opposite side of the circle.

To check on the possibility that these results were peculiar to one set of pictures, the method was applied to two other series. In an unpublished Honors project in 1943, Miss Marjory L. Brown found that the Ruckmick (**5**) series showed clear evidence of conforming to a circular scale, with five of the 32 pictures having some overlap between Steps 6 and 1. Another series was also tried, made up of supposedly unposed pictures. Fifteen were from the Munn (**4**) series, and 17 additional ones were selected from *Life* magazine. Only one of this series spread across the "ends" of the scale, but this partial failure to show circularity may have been due to the selection of pictures. There was only one with a mode in Step 6 (Contempt) and the only four pictures in Step 1 showed very little spread to any other

categories. Hanawalt (2) has also reported a failure to find evidence of a circular scale in facial expressions, but this finding was only an incidental one in an experiment designed for other purposes.

The question as to whether or not the Woodworth scale of facial expressions constitutes a circular or recurrent series is really only a preliminary one. Let us assume for the moment that the scale describes the periphery of a roughly circular surface on which all facial expressions can be located, and examine the consequences. The circular surface may be compared with the color circle, with the six scale steps corresponding to the major hues. This comparison brings out the fact that the facial expression surface should have a neutral point ("gray") in the center; the strength of an expression would correspond roughly to the saturation of a color. Still more important, one should be able to describe the surface in terms of two axes, just as the color surface can be described in terms of blue-yellow and red-green axes. Examination of the results of the 1941 experiment (6) led the writer to suggest that these two axes were pleasantness-unpleasantness and attention-rejection. The first axis needs no further explanation, but the second one does. Attention is exemplified by surprise, in which all receptors are maximally open to stimulation. Rejection is the best term we have found for the other end of this axis; it is shown most clearly in contempt and disgust, in which eyes and nostrils appear to be actively shutting out stimulation (Fig. 2).

This brings us to the really important question—can typical series of facial expressions be adequately described in terms of two axes? If we can answer the question in the affirmative, a very complex field will have been reduced to relatively simple dimensions. One way to test this hypothesis is to obtain ratings of the locations of each of a series of pictures on the P–U and A–R axes, and then to use this joint position to predict the Woodworth scale value of each picture. The present paper reports four separate studies. We may anticipate the results by pointing out that three of the studies, using two different series of pictures, yielded correlations above .90 between obtained scale values and those predicted from independent ratings of P–U and A–R.

Procedure

All four experiments employed the same basic method. The *S* sorted a set of pictures along a 9-point rating scale ranging from maximum unpleasantness (1) to pleasantness (9). While *E* was tabulating these ratings, *S* was given duplicate prints of the same pictures to sort on a scale rejection (1) to attention (9). Ratings from all *S*s were combined by simple averaging to give a single pair of values for each picture. The pictures were then plotted on a sheet of graph paper, using P–U values on the ordinate and A–R values on the base line. This scatter plot was then mounted on a large 360° protractor, with the center of the circle at the mid-point (5–5) of the axes, and the P–U axis oriented at 60° and 240°, corresponding to scale positions 1.00 and 4.00, as we will see in a moment. A thread stretched from the center, across a plotted point and out to the periphery, permitted a reading for that picture in degrees, referred to the circumference of the circular surface. The degree reading could be converted to predicted positions on the Woodworth circular scale by dividing by 60, since there were six steps on the Woodworth scale. These predicted values were then compared with those actually obtained by independent sorting made in the Woodworth bins.[1]

[1] If the method of predicting Woodworth scale position from rating data is not clear from this brief description, turn to Fig. 1, under *Results*. In the Discussion, it will be shown that this method unavoidably discards some information. The location of a picture on a surface described by two axes can give (*a*) the distance the picture lies from an arbitrary center, and (*b*) the radius on which it falls. Since the Woodworth scale is simply the circumference of the circular surface, we can utilize only (*b*), above, in making the predictions.

TABLE 1

PREDICTED AND OBTAINED SCALE VALUES FOR THE RUCKMICK PICTURES

Description*	Picture Number	P–U	A–R	6-Step Scale Position			1943 Scale Data	
				Predicted	Obtained	Error	Position	Av. Dev.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reconciliation	28	8.05	7.15	1.60	0.90	0.70	0.92	0.19
Interested Inquiry	30	7.90	7.10	1.62	0.92	.70	1.01	.05
Amusement	29	8.20	7.25	1.60	1.00	.60	1.00	.00
Mirth	31	8.80	8.15	1.67	1.00	.67	1.00	.00
Joy	32	8.80	8.25	1.68	1.00	.68	1.00	.00
Adoration	17	6.60	6.25	1.65	1.05	.60	1.20	.29
Wistful Appeal	26	6.95	6.75	1.70	1.22	.48	1.10	.20
Cynical Interest	19	6.70	6.40	1.65	1.23	.42	1.41	.41
Interested Observation	27	7.20	6.85	1.67	1.27	.40	1.28	.45
Astonishment	23	4.60	5.95	2.87	2.05	.82	2.10	.28
Consternation	24	3.85	7.10	2.98	2.05	.93	2.10	.30
Pleading	20	5.25	5.70	2.17	2.08	.09	2.50	.87
Apprehension	2	2.25	6.10	3.65	2.70	.95	2.77	.59
Startled Fear	3	1.75	5.30	3.92	2.72	1.20	2.67	.50
Dread	7	2.05	5.75	3.77	2.85	.92	3.00	.30
Anguish	8	2.05	5.10	3.97	2.97	1.00	3.00	.08
Resentment	14	3.30	4.90	4.09	3.13	.96	3.45	.75
Interested Attention	18	3.25	6.45	3.35	3.15	.20	2.59	1.06
Sorrow	5	4.65	5.50	3.12	3.19	−.07	2.85	.40
Exhausting Pain	6	2.40	3.15	4.60	3.47	1.13	3.75	1.13
Anxiety	21	3.40	4.95	4.05	3.63	.42	3.28	.52
Fear	4	1.45	5.90	3.77	4.00	−.23	3.93	.54
Anger	1	2.00	4.60	4.13	4.17	−.04	4.13	.38
Distrust	12	2.65	4.20	4.32	4.37	−.05	4.65	.69
Haughtiness	15	4.35	3.75	5.05	4.52	.53	4.72	.85
Contemplation	10	3.70	3.75	4.75	4.62	.13	4.36	.51
Defiance	16	3.10	3.80	4.55	4.73	−.18	4.50	.77
Critical Distrust	11	3.50	3.45	4.79	4.87	−.08	5.14	.79
Rueful Meditation	22	4.55	3.70	5.19	5.27	−.08	4.74	.91
Scorn	13	3.60	2.25	5.05	5.58	−.53	5.68	.51
Sulkiness	9	4.00	2.60	5.13	5.68	−.55	3.50	.80
Contempt	25	6.00	5.55	7.47	6.40	1.07	6.10	.33
Mean		4.59	5.42	(r = .96)		.54		

* The descriptions are slightly condensed from the "Adopted Description" column of Ruckmick's leaflet which is furnished with the set of photographs by C. H. Stoelting Co., Chicago, Ill. The numbers used in this study are those printed on the backs of the photographs; they do not correspond to those given in Ruckmick's original paper (**5**). The pictures are listed in order of Woodworth's scale position (Col. 5), with breaks to separate the six steps.

In Exp. I, the Frois-Wittmann pictures were used. The P–U and A–R sorting was done in piles on a series of nine numbered cards. Little trouble was found in getting *S* to understand the P–U axis; it was merely necessary to stress the fact that he was to judge whether "The man felt pleasantness or unpleasantness." The A–R axis caused more trouble. An effort was made to describe attention as *openness to stimulation* and to point out that the eyes, nostrils, and mouth tended to be constricted in rejection.

In Exp. II, a more effective way was found for describing the rating scales. After the usual explanation of A–R, picture No. 23 of the Ruckmick series was shown as a typical example of attention, and No. 13 as rejection. After *S* had placed these pictures at the appropriate positions on the scale, these "anchors" were reclaimed, and he was given the pack of Frois-Wittmann pictures to sort. The *S* was allowed to review the anchors if he wished to do so.

The results of Exp. II were so encouraging that it seemed desirable to duplicate this experi-

ment as a check. Experiment III was run a year later, with only minor changes over II. Two wooden racks of bins were used to facilitate handling and tabulating the pictures, and fresh prints of the pictures were used.

Experiment IV was planned to check the results of II and III on another series, that of Ruckmick, which consisted of 32 pictures of a female face. Both the P–U and A–R scales were anchored by showing pictures from the Frois-Wittmann series, each end being exemplified by six pictures. In this experiment, *Ss* sorted the pictures into the Woodworth scale bins both before and after the rating, using the same rack and method employed earlier (**6**). This was done largely to build up a new set of scale values for this series, which could be compared with those obtained approximately a decade earlier.

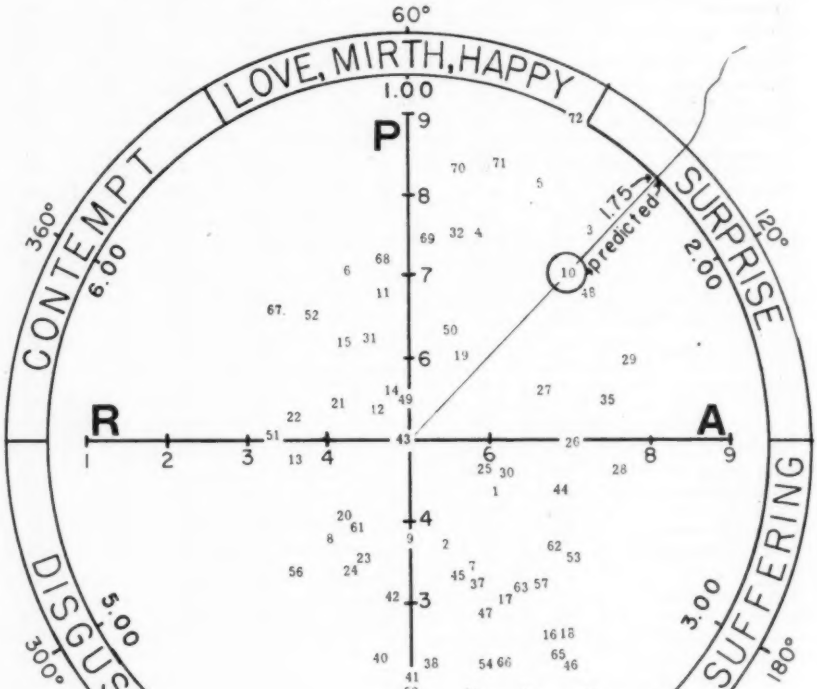

FIG. 1. The location of each of the 72 Frois-Wittmann pictures on an oval surface. Picture No. 10 (encircled) serves as an example of the method used. It is plotted at axis values of P–U = 7, A–R = 7, as determined by the rating scales. A thread is then stretched from the intersection of the axes (5–5) across the plotted point, and its position read off in degrees. Dividing by 60 yields the predicted Woodworth scale value (1.75) which is compared with that obtained earlier by direct sorting on the Woodworth scale (1.65). This picture might be called, "Pleased Surprise," and both predicted and obtained values place it in the appropriate step of the Woodworth scale.

TABLE 2

CORRELATIONS AND AVERAGE ERRORS BETWEEN PREDICTED AND OBTAINED SCALE VALUES

Exp.	Pictures Used	Type of Anchor	*Ss*	*r*	Average/Error
I	Frois-Wittmann	Verbal	8	.76	not avail.
II	Frois-Wittmann	Samples	9	.94	not avail.
III	Frois-Wittmann	Samples	18	.92	.52, .46*
IV	Ruckmick	Samples	20	.96	.54, .33*

* These average errors were obtained against an adjusted scale.

(Experiments I–III used the Schlosberg scale values for the Frois-Wittmann pictures.)

All experiments were conducted by undergraduates as special projects in the writer's laboratory course.[2] The students made their own analyses of the data, but the writer has verified and extended all computations in Exp. III and IV. All *Ss* in the first three experiments were members of the same course, and were approximately equally divided as to sex. There were eight *Ss* in Exp. I and nine in II. In Exp. III, there were 18, but two *Ss* repeated the experiment, giving 20 sets of data. Of the 20 *Ss* in Exp. IV, 8 were from the laboratory course, and the remaining 12 were nonstudents, ranging in age from 20 to 55 yr. No *S* participated in more than one experiment.

RESULTS

Table 1 presents the detailed results from the Ruckmick series in Exp. IV.[3] For comparison with the present data, the unpublished results obtained in 1943 by Miss Brown are also presented. The older scale positions were based on three sortings by each of 23 *Ss*. The method of deriving the predicted scale position from P–U and A–R ratings is given in the Procedure, and is illustrated in Fig. 1.

[2] I wish to express my thanks to the students who served as *Es* in the following experiments: I—Robert F. Shepard, Jr.; II—Mark T. Sheehan; III—Louise Anthony, Jean M. Coraci, Sheila K. Hart, Ann M. Thomas; IV—Shirley A. Juskalian, John J. Kennedy. Dr. W. S. Hulin kindly supplied several sets of the Frois-Wittmann pictures, which speeded up the collection of data.

[3] The data on individual pictures of Exp. III will not be tabulated, since the ratings and predicted scale positions can be determined from Fig. 1 and the obtained scale positions are in print (6).

Table 2 summarizes the results of all four experiments together with key words to identify the methods. Notice that all three experiments which employed anchors yielded predictions which correlated above .90 with the scale positions obtained by sorting the pictures into the Woodworth scale. One may question the precise significance of these correlations, for the scale values are not normally distributed. Further, the coefficient does not yield all the information that might be desired.

A more revealing analysis can be made by examining the errors in prediction. Mean errors were computed for Exp. III and IV, and are included in Table 2. In both experiments, the average error of the prediction was about half a scale step; chance would give three times this error, for in a circular scale of six steps the maximum error is three steps, and the average error would be one and one-half steps.

DISCUSSION

The correlations and the relatively small errors show that reasonably good predictions of Woodworth scale positions can be made from ratings on two axes, pleasantness-unpleasantness and attention-rejection. There is reason to believe that a large portion of what errors there are may be attributed to the scale itself, and to certain working assumptions made in relating the scale to the axes. These assumptions were (*a*) that the center of the

circular scale was at axis position 5–5, the mid-points of the two 9-point rating scales, (*b*) that axis ends P, A, U, and R fell at circular scale positions 1.0, 2.5, 4.0, and 5.5, respectively, and (*c*) that the six steps on the circular scale were equal. The results should enable us to refine these assumptions, and to improve the prediction.

Examination of Table 1 will show that the second assumption is markedly in need of correction for the Ruckmick series. Column 6 shows a marked preponderance of positive errors, in which the axis positions predicted scale values that were too high. The constant error is approximately .43 scale steps; when most of this is removed by rotating the plotted surface counterclockwise, putting P at .60 scale position, the average error drops from .54 to .42 scale steps. Simple rotation was not tried on the Frois-Wittmann results (Exp. III) for there was little indication of a constant error throughout the scale.

Assumption (*a*), the locus of the mid-point of the circular scale, does not seem to be critical. Recalculations were made on both series of pictures after shifting the center to axis positions 5–5¼; 4–5; and on the Ruckmick series, 4¼–4¼. These shifts increased the average error slightly, but never beyond one of .55 scale points.

Assumption (*c*), that the units are equal throughout the scale, is obviously only a gross approximation. In the Ruckmick series the variability of judgment, as shown by the average deviation, is about one-third as large in Step 1 as in the other five steps. This suggests that Step 1 is three times as big as any of the other steps. As a way of testing the effects of a rough adjustment of the circular scale, a base line was marked off with the first step doubled in width, and the others cut to .8 units. Using predicted scale positions on the ordinate, the Ruckmick circular scale positions were plotted on this base line, and a line with a slope of one was fitted to the points by inspection. The deviations of the predictions from this line were then measured and averaged, yielding a mean error of .33 scale steps. This was by far the smallest average error of prediction obtained. A similar treatment of the Frois-Wittmann data, using a scale adjusted to conform to the average deviations of this series (**6**), gave an average error of .49 scale steps. Omitting one picture which cannot be adequately located on the circular scale (No. 43, see below), the average error becomes .46.

An average error of prediction of .33 scale steps is probably close to the minimum. One limiting factor is the stability of the circular-scale positions we are trying to predict. In Table 1, two sets of such positions are given for the Ruckmick series, the set obtained in Exp. IV, and that obtained a decade earlier. The average difference between these two sets of values is .23. The reason for this difference is not hard to find; the scale forces all pictures out to the circumference, when some of them belong near the center of a circular surface. This is particularly clear in the case of picture No. 43, which appears at the center of both Fig. 1 and 2. It received an axis rating of 5.0 and 4.9, leading to a predicted position of 5.50 on the circular scale. But if 4 of the 20 *Ss* had increased their rating by one step on the A–R scale, thereby changing the picture's mean rating to 5.1 on the other side of the center, the prediction would have been 2.50. Thus the circular-scale position of this picture is largely a matter of chance. This showed clearly in the results of the 1941 experiment, for about one-third of the *Ss* put this picture in the "Scattering" bin, and the rest distributed it almost randomly in five of the six main categories. This picture is an extreme case, but there were others that showed large spread; for example, 8 of the 32 Ruckmick pictures showed average deviations (Table 1, col. 8) greater than .75 scale steps. For such pictures it seems probable that the axis positions give a better description than do the circular-scale values. That is, a relatively neutral picture like No. 43 belongs not at the circumference of the circular surface, but near its center.

Now that we know we are dealing with the whole surface rather than with its circumference, we can ask for a description of the surface. Like the "color circle," it is only roughly circular. Figure 2 probably gives a reasonable first approximation to its outline, for it includes all the extreme pictures. It was not practical to mount all the pictures in the more crowded areas, but their location is indicated by the scatter diagram reproduced in Fig. 1. The surface seems to be roughly oval in form, with the P–U axis longer than the A–R axis. This agrees with the conclusion drawn in the writer's 1941 paper, based on the distribution and average deviations of the pictures on the circular scale. It is difficult to determine the exact form of the surface, for there is no reason to believe that the distances

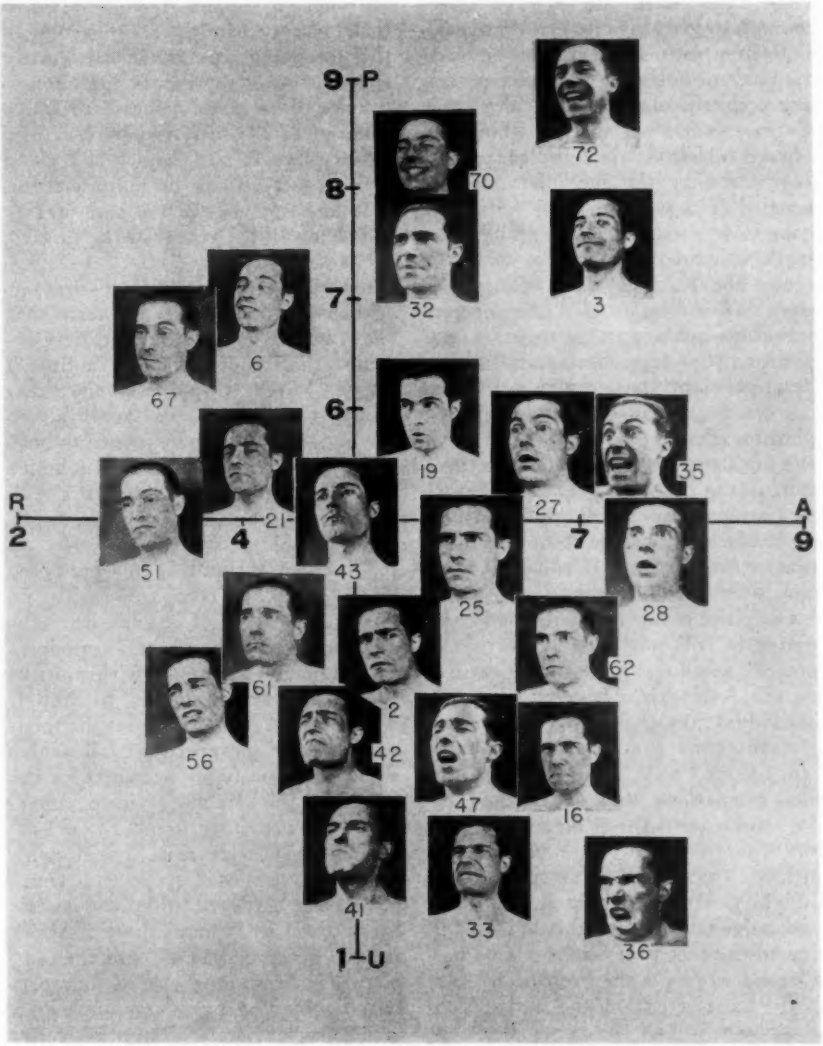

FIG. 2. The location of typical pictures of the Frois-Wittmann series. This display includes those pictures which are shown around the margin of the surface in Fig. 1, and as many of the more centrally located ones as could be mounted. The average error of prediction for these 25 pictures is a half-scale step, just as it is for the whole series of 72 pictures. The picture with the worst prediction of the whole series is No. 43, at the intersection of the axes (see text).

are equal on the two axes. The present rating data are not extensive enough to warrant any elaborate scaling attempts. There is also the question as to whether or not the two axes are perpendicular; the difficulty some *Ss* experienced in distinguishing between rejection and unpleasantness suggests that the included angle is somewhat less than 90°. Perhaps some of these relations might be worked out by comparison with the Hulin and Katz (**3**) data. They had their *Ss* sort the Frois-Wittmann pictures into piles, putting together any pictures that seemed similar. If our description of the circular surface is correct, the distance between any two pictures should vary inversely with the number of times the two pictures were put in the same pile. But before this comparison is attempted, it would probably be advisable to obtain more precise ratings on P–U and A–R than our 20 *Ss* have furnished.

One final point may be added—the writer is still not satisfied with the present method of obtaining ratings on the A–R axis. Inspection of the individual ratings assigned to the same picture shows that they occasionally range from 1 to 9. The fact that the final predictions work as well as they do shows that these errors average out; there seems to be a genuine dimension that shows up in spite of the difficulty with which it is described and judged. Perhaps the difficulty is due to the fact that our language has no good word for the "rejection" end of the axis; attention and inattention describe only half the axis. But certainly the fact that common speech is not cognizant of the whole range of a dimension proves little about the utility of that dimension in analyzing facial expression, except only that the dimension may be difficult to handle experimentally or to describe in any but operational terms.

The question was raised earlier as to whether or not the Woodworth scale was recurrent. The results of the foregoing analysis show clearly that the scale represents the circumference of an oval surface into which some pictures are forced by lack of a better place to put them. If a given series contains one or more pictures that lie somewhat above and slightly to the left of the center of the surface, as at axis position P–U 6, A–R 4.5, they will be projected to a circular-scale position of 6.5. This means that they will be judged in Step 6 (Contempt) half the time, and in Step 1 (Love, Mirth, Happiness) the other half of the time. This spread across the ends of the scale happened in both the Frois-Wittmann and Ruckmick series, and furnished the original evidence for circularity. For example, the Frois-Wittmann picture No. 15 fell at 6.2–4.2, resulting in a predicted value of 6.21, compared with an obtained value of 6.27. But the obtained value represents a mode of 54 judgments in Step 6 with most of the spread, 50 judgments, to Step 1. The Frois-Wittmann series happens to be rich in pictures that fall in this area, which probably accounts for the fact that the circularity is so apparent. On the other hand, there happened to be no picture in Miss Brown's supposedly unposed series that could have been expected to spread across Steps 6 and 1. The pictures of Mirth showed no spread at all, and the only picture in Contempt was very close to Disgust, leaving a gap of a whole step at the crucial point. Parenthetically, it may be added that Contempt is a sophisticated expression, and is rather hard to judge. It is interesting that the oval surface has a place on which it fits so nicely.

But to return to the circularity of the Woodworth scale. It was a fortunate choice of pictures that led to

the original detection of the circularity, for there is no assurance that a given set of pictures will show the overlapping spread across the two supposed ends of the Woodworth scale. Failure to find such spread does not alter the more fundamental fact that all pictures of two different series, 72 posed by a man and 32 by a woman, can be placed rather well on an oval surface described in terms of two axes.[4] These axes are proposed as the basic dimensions in which facial expressions may vary. The fine shades that are familiar in literary usage may well involve additional factors, as knowledge of the situation and of the past history of the individual.

Summary and Conclusions

1. In three independent experiments, using 8, 9, and 18 *Ss* respectively, ratings of the 72 Frois-Wittmann pictures of facial expressions were obtained on two nine-point rating scales: Pleasantness-Unpleasantness and Attention-Rejection. In a fourth experiment with 20 *Ss*, similar ratings were obtained on the 32 Ruckmick pictures.

2. Using these rating scales as rectangular axes, each picture was located on a roughly circular surface.

3. By referring the plotted position of each picture to the circumference of a circle drawn around the surface, an attempt was made to predict the value each picture had received on a recurrent six-category scale of the Woodworth type. When the predictions were based on a poorly-anchored A–R scale (Exp. I), they yielded a correlation of .76 with obtained values. Anchoring the rating scales by samples from another series of pictures (Exp. II, III, IV) raised the correlation to .94, 92, and .96, respectively.

4. The average error of prediction was of the order of half a scale step, which is one-third the error expected on a random series on a six-step circular scale. By adjustment of axes and scale steps, the error of prediction could be made to approach the limits set by the accuracy of the original scale judgments.

5. It is concluded that two widely used series of facial expressions can be described very well by locating them on a roughly oval surface whose longer axis is pleasantness-unpleasantness, the shorter axis being attention-rejection. Recognition of finer shades of emotion may depend on knowledge of the stimulus situation.

(Received February 25, 1952)

[4] Many of our *Ss* reported that they tended to judge P–U on the basis of the mouth, and A–R from the eyes. This is consistent with Hanawalt's (2) report that happiness is often confused with pain-suffering when judgments are based on only the upper half of unposed pictures. If the mouth is blotted out, *S* should have difficulty in distinguishing between P and U, the dimension which differentiates happiness from pain-suffering. Careful study of confusions in half-face judgments might throw further light on the determiners of our two axes.

References

1. Frois-Wittmann, J. The judgment of facial expression. *J. exp. Psychol.*, 1930, **13**, 113–151.
2. Hanawalt, N. G. The role of the upper and lower parts of the face as a basis for judging facial expressions: II. In posed and "candid-camera" pictures. *J. gen. Psychol.*, 1944, **31**, 23–36.
3. Hulin, W. S., & Katz, D. The Frois-Wittmann pictures of facial expression. *J. exp. Psychol.*, 1935, **18**, 482–498.
4. Munn, N. L. The effect of knowledge of the situation upon judgment of emotion from facial expressions. *J. abnorm. soc. Psychol.*, 1940, **35**, 324–338.
5. Ruckmick, C. A. A preliminary study of the emotions. *Psychol. Monogr.*, 1921, **30**, No. 3 (Whole No. 136), 30–35.
6. Schlosberg, H. A scale for the judgment of facial expressions. *J. exp. Psychol.*, 1941, **29**, 497–510.
7. Woodworth, R. S. *Experimental psychology*. New York: Henry Holt, 1938.

REMOTE ACTION POTENTIALS AT THE MOMENT OF RESPONSE IN A SIMPLE REACTION-TIME SITUATION [1]

ROBERT L. HENDERSON

University of Missouri

Meyer (6) has reported an experiment designed to test Dunlap's hypothesis that practice in a simple reaction-time (RT) situation will result in progressive decrease of muscular contractions in nonparticipating body parts at the moment of each reaction. Meyer's results failed to provide any confirmation of this hypothesis, and he concluded that the presentation of the simple reaction-time task may give rise to either positive or negative increments of amplitude of remote muscular contractions. While admitting that his data are influenced by a number of incompletely controlled variables, Meyer suggests that the Dunlap hypothesis requires revision (6, 7).

Daniel (2) has subsequently pointed out certain other weaknesses in the design and execution of the Meyer study which leave the validity of the published results questionable. The present study represents an effort to test further Dunlap's original hypothesis, using techniques and procedures similar to those used by Meyer, while trying to correct certain of his errors.

Experimental Design

Subjects.—The *S*s were 20 undergraduate students in psychology at the University of Missouri. All *S*s were obtained on a volunteer basis to insure adequate motivation, and were naive as to the purpose of the experiment.

Action potential apparatus.—The *S*s were isolated from the electronic equipment by a shielded cage constructed of copper screen wire properly grounded. Shielded leads from the surface electrodes connected into one channel of a four-channel Rahm amplifier designed for bioelectric measures. The output of this amplifier was fed directly onto the vertical plates of a cathode ray oscilloscope, with the resulting signal photographed by a 35-mm. moving film camera.

Power was applied to the electronic apparatus at least 1 hr. before experimentation was to take place to insure proper "warm-up" of the equipment. At the beginning and end of each session, the apparatus was calibrated. This was accomplished by feeding a 100-cps tone from an oscillator, through a microvolter, into the input of the amplifier. By systematically varying the input voltage and measuring the corresponding amplitude of the photographed signal, it was possible to graph the output amplitude as a function of input microvoltage.

Reaction time apparatus.—The reaction studied was that of a simple key-pressing response to a visual stimulus. A standard telegraph key was mounted on a table before *S*. A 1-in. diameter round window was cut in a cardboard screen and covered with translucent white paper. The screen was then mounted so that the stimulus window was at *S*'s eye level and at reading distance.

A small slide projector located outside the shielded cage served as the light source. It was directed in such a manner that a small beam of light illuminated both the stimulus window and a small mirror mounted on *S*'s key. A magnetic shutter, normally closed, was used to block the beam of light between trials. By pressing a key, *E* could open the shutter and simultaneously illuminate the stimulus window and the recording mirror. This mirror reflected the beam of light to a second mirror, mounted on the face of the oscilloscope which in turn reflected the light into the camera where it was photographed as a solid black line. The slightest depression of *S*'s key deflected the beam of light so that it was no longer photographed.

At the opposite side of the oscilloscope from the RT mirror was a second light source which was interrupted 100 times per second by a circular shutter driven by a constant speed motor. On the photographic record this appeared as a

[1] Graduate research carried out at the University of Missouri under the direction of Dr. Robert S. Daniel, to whom the writer is gratefully indebted for suggesting the research and assisting in its design and execution.

series of discrete dots, each representing a time interval of .01 sec.

Procedure.—A major departure from the Meyer design was the use of the left arm instead of the right leg for purposes of obtaining action-potential records. Daniel (1) had shown earlier that the action potentials in the left arm are stronger than those in either leg when the right arm is active. Preliminary tests of the apparatus and procedure in the present study confirmed Daniel's results and justified the change in procedure. It should be noted that while this change prevents a direct comparison of the results of the present study with those obtained by Meyer, it is not inconsistent with the hypothesis as originally stated by Dunlap.

A second important departure from the procedure used by Meyer was the adjustment of the amplifier gain control during each session to maintain as large an action-potential amplitude as could be accommodated by the 3-in. cathode ray tube of the oscilloscope. Had Meyer's procedure of maintaining the same amplifier gain throughout the experiment been followed, it is very probable that a number of responses would have been lost inasmuch as it was frequently necessary to increase the sensitivity of the amplifier during the course of each session to maintain a measurable action-potential signal.

Upon entering the experimental room on the first practice day, *S* was seated before the RT key and given a set of printed instructions to read. These instructions were as nearly identical to those used by Meyer as the changes in experimental design permitted.

The active electrode was then placed over the thickest portion of the *flexor carpi radialis* and the neutral electrode over the thickest portion of the upper arm above the unused hand. After answering any questions on procedure asked by *S*, *E* left the shielded cage and proceeded to the AP apparatus, out of sight of *S*.

At approximately 10- to 15-sec. intervals *E* would say "ready," and, after an interval of from 2 to 4 sec., would press the key actuating the stimulus light shutter. The shutter then remained open until *E* either saw or heard the *S*'s reaction.

Each of the 20 *S*s were given six practice periods, one every two days. Each period consisted of 100 responses, with every tenth response recorded photographically. The average of the ten recorded responses was used as *S*'s score on that day. A list containing each *S*'s name and his individual mean RT score for each practice day was kept in the experimental room. As in Meyer's study, each *S* was allowed to inspect this list before each practice session and compare his own score with that of the other *S*s.

The throwing of the clutch connecting the camera drive to a continuously running electric motor introduced a slight change in noise level and tone. However, introspective reports obtained from *S*s at the end of the experiment indicated that none realized that only every tenth response was being recorded. It should be mentioned that the camera drive motor and its associated reduction and drive gears provided a noise level more than adequate to mask the click of the magnetic shutter controlling the stimulus light, eliminating the possibility that *S*s learned through practice to respond to the auditory stimulus of the shutter rather than to the stimulus light.

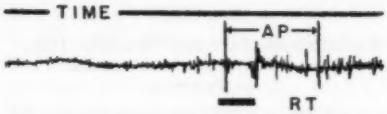

FIG. 1. A typical record showing the time line, with each dot equal to .01 sec., the action potential with the measured area indicated, and the reaction-time line indicating the time interval from the onset of the stimulus to *S*'s reaction

Measurement.—The sample appearing in Fig. 1 indicates that typical action-potential records were obtained. Inasmuch as the response in question appeared in the form of a group or cluster of spikes varying in number and temporal association, a measure of the spike of greatest amplitude was not considered to give a comprehensive picture of the entire response. Consequently a procedure similar to that used by Davis (3, 4) was utilized. This consisted of summating microvoltages of the three largest spikes which occurred within a period of .1 sec. before the response and .2 sec. following the response. Each *S*'s action-potential score (AP) for each day was, then, the mean of these summated values for the ten responses recorded.

The RT score was obtained very simply by counting the number of .01-sec. dots which occurred within the period from the onset of the stimulus until the reaction was made.

RESULTS

Table 1 contains the mean RT for all 20 *S*s on each of six successive practice days along with the mean AP in microvolts. Except for a small reversal on Day 4, it will be seen that the mean RT decreases steadily, while the mean AP drops consistently for the first four days, then rises on Days 5 and 6.

TABLE 1

RELATIONSHIP BETWEEN REACTION TIME, ACTION POTENTIAL, AND AMOUNT OF PRACTICE

Practice Day	RT in .01 Sec.		AP in Microvolts	
	Mean	*SD*	Mean	*SD*
1	21.32	1.51	54.78	23.45
2	18.61	1.91	53.16	26.11
3	17.65	1.94	38.78	18.17
4	17.90	2.55	33.06	14.44
5	17.08	1.91	40.64	19.66
6	16.75	1.89	40.65	23.59

After utilizing Bartlett's test for homogeneity of variance, a trend analysis was run on the AP data using the analysis of variance technique. This is summarized in Table 2. The obtained F of 4.76 (for 5 and 95 *df*) is significant at the 1% level of confidence and indicates that the curve of the AP means, shown in Fig. 2, differs significantly from a horizontal straight line.

Further analysis with the t test indicates a significant difference between Day 1 and Day 4 ($t = 3.05$; 2.86 needed at 1% level) and between Day 1 and Day 6 ($t = 2.13$; 2.09 needed at 5% level), but no significant difference between Day 4 and Day 6 ($t = 1.46$; 2.09 needed at 5% level). Hence the terminal rise in AP may be attributed to chance fluctuation and is not necessarily indicative of a real rise.

A second analysis of the data was run on the tension level during the foreperiod. These data are, in general, consistent with those obtained on the AP itself and will not be reported here in detail. The same trends were present with the mean tension level decreasing for the first four practice days and rising again on the last two days. A trend analysis following Lindquist's method (5) indicated that the two curves were essentially parallel, with the curve for the AP data significantly higher than the curve for the foreperiod tension.

TABLE 2

TREND ANALYSIS OF AP DATA

Source	*df*	Sum of Squares	*F**
Total	119	61645.05	
B_I	5	7358.49	4.76
B_S	19	24889.08	
$I_I \times S$	95	29397.48	

* $F = 4.76$ is significant at the 1% level of confidence.

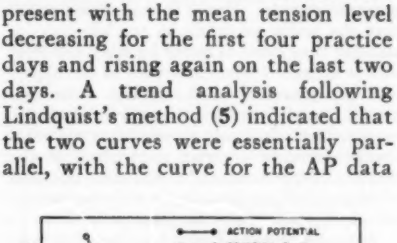

FIG. 2. Relationship between mean reaction time and mean action potential on each of six successive practice days

DISCUSSION

The results of the present study are consistent with the hypothesis that tensions in a nonparticipating body member declines during learning or practice of a response by a restricted muscle group. The level of action potential is seen to decline steadily during early learning of a simple reaction-time response, then to rise as the practice is continued. Although the terminal rise in action-potential level was not statistically significant over the period covered by the present investigation, the trend obtained suggests that had practice been continued the rise might have attained statistical significance.

A tentative hypothesis may be

offered to explain such a terminal rise in tension level. We may note that the rise in tension coincides with the asymptote of learning of the response in question. It seems likely that the *Ss*, under high motivation in a competitive situation, would tend to become more "anxious" or more "tense" when they find, late in practice, that their reaction-time scores are showing less and less improvement from session to session. Such an increase in generalized tension would be expected to raise the absolute value for the action potential inasmuch as it is added to, or superimposed upon, the general tension level.

This hypothesis would suggest that a more adequate test of the Dunlap hypothesis would consist in partialling out the tension present in *S* prior to the response or the onset of the preparatory set, and analyzing the remaining action-potential values which may be considered more or less directly to be a function of the response itself.

Summary

Action-potential measurements were taken from a nonparticipating body member of a group of 20 *Ss* responding in a simple reaction-time situation to a visual stimulus. The reaction time was found to decline steadily over a period of six successive practice days of 100 trials each. The action potential during the foreperiod and at the moment of the response was found to decline during the first four days of practice, then to rise on Days 5 and 6. Although the terminal rise in action potential was not statistically significant, a tentative hypothesis was advanced which explains the rise in terms of increased general tension arising from increased motivation to improve as the asymptote of learning is reached.

(Received February 20, 1952)

References

1. Daniel, R. S. The distribution of muscular action potentials during maze learning. *J. exp. Psychol.*, 1939, **24**, 621–629.
2. Daniel, R. S. Some observations on Meyer's study of reaction time and muscle tension. *J. exp. Psychol.*, 1949, **39**, 896–898.
3. Davis, R. C. Set and muscular tension. *Indiana Univer. Publi. Sci. Ser.*, 1940, No. 10.
4. Davis, R. C. Methods of measuring muscular tensions. *Psychol. Bull.*, 1942, **39**, 329–346.
5. Lindquist, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65–78.
6. Meyer, H. D. Reaction time as related to tensions in muscles not essential in the reaction. *J. exp. Psychol.*, 1949, **39**, 96–113.
7. Meyer, H. D. Some remarks concerning Daniel's observations. *J. exp. Psychol.*, 1949, **39**, 898–900.

DIFFERENT STRENGTHS OF SET IN THE SOLUTION OF ANAGRAMS [1]

IRVING MALTZMAN AND LLOYD MORRISETT, JR.

University of California at Los Angeles

A major problem confronting the investigator of human problem solving has been to account for the "direction of thought," the sequence of "ideas." In a problem situation one relevant attempt at a solution occurs out of a great number of possibilities. It may be rejected if incorrect, retained if correct. If incorrect, other relevant attempts may occur, while still others do not occur at all. Investigators in this area have tried to account for the direction of thought with the aid of such concepts as determining tendency, Aufgabe, set, direction, stress, and so on (5).[2]

There have been numerous experimental demonstrations of the operation of mental set in problem solving (2). One such study related to the present experiment is that of Rees and Israel (8). The problems used by these investigators were five-letter anagrams. In one experiment *Ss* were required to solve anagrams that had only one possible solution, either a word referring to some aspect of nature or a word referring to some aspect of eating. The instructions given did not in themselves lead to the arousal of a particular set for one or the other of these classes of anagrams. Following training on either the eating-set or nature-set anagrams, a test series was given in which the anagrams had several possible solutions only one of which was a nature or an eating word, depending upon the previous training conditions. The higher percentages of *Ss* with set training that gave eating or nature solutions on the test series as compared with *Ss* that did not have such training indicated the establishment of a set.

From the point of view of S–R behavior theory the problem of the direction of thought is similar to the problem confronting the learning theorist attempting to account for the course of trial and error or selective learning. Hull has attempted to do this in terms of principles derived from conditioning experiments (3). The basic assumption of this paper is that these same principles can be employed in accounting for the direction of human problem solving.

Hull (4) has assumed that habit strength increases as a negatively accelerated function of the number of reinforcements, specifically, as an exponential function. It follows that with repeated trials in a choice situation a difference in habit strength would develop more rapidly between two weak habits of equal strength than two strong habits of equal strength.

Siegel (9) tested this implication by running two groups of rats in a T maze. One group received 40 reinforcements in the left arm and 40 reinforcements in the right arm. The second group received two reinforcements in the left and two reinforcements in the right arm. The *Ss* of both groups were then given 48 trials

[1] This study was supported by a grant from the Committee on Research, University of California.

[2] This paper will be concerned with the problem of mental set as distinguished from direction. The distinction is similar to that made by Maier (6), although it would differ in the details of the related theory.

in which reward was received in only one arm of the T maze, either the left or the right. It was found that the weak habit-strength group making 43 mean correct responses learned significantly faster than the strong habit-strength group making 39 mean correct responses. This is in accord with the implication of Hull's theory.

The design of the present experiment is similar to Siegel's in its establishment of weak and strong competing reaction potentials. The problem situation was the simple anagram having only one correct solution (8). Competing response tendencies were the class of response solutions elicitable by the class of eating anagrams and the class of response solutions elicitable by the class of nature anagrams.

The hypothesis tested in the experiment to be reported is that if "mental set" refers to the acquisition of habit strength by a class of stimuli for the elicitation of a particular class of responses, the implication of Hull's theory for different absolute initial habit strengths should also hold for incompatible anagram sets. There should be an inverse relationship between amount of practice with incompatible classes of anagrams and subsequent success with one of them.

Method

Subjects.—Forty *S*s from introductory psychology classes were employed in the experiment, 20 in each of the two major conditions.

Stimulus materials.—Five-letter anagrams typed in lower case on white 3 × 5 in. index cards were used as the problems. Most of the anagrams were obtained from lists published by Rees and Israel (8). Some of these had to be discarded following preliminary experimentation when alternative solutions belonging to different classes were discovered. Additional anagrams were constructed by the writers. The anagrams finally used in this study are presented in Table 1.

Three different classes of anagrams were used, eating-set, nature-set, and neutral anagrams. Most of the eating- and nature-set anagrams had only one possible solution, a word referring to some aspect of nature or eating. In the few cases where alternative solutions were possible, the alternative belonged to the same class of responses. Neutral anagrams had one or two solutions neither of which was directly or obviously related to eating or nature.

Procedure.—The strong habit-strength group received 15 eating- and 15 nature-set anagrams in a prearranged order based on drawings from a table of random numbers. The weak habit-strength group received 26 neutral anagrams

TABLE 1

Anagrams Employed in the Experiment

Practice		*Nature-set Training*	
cikbr	brick	lutpi	tulip
ihrac	chair	rokbo	brook
saree	erase	sarts	stars
		otsmr	storm
Neutral Training		edses	seeds
soobk	books	esnfr	ferns
labls	balls	cheab	beach
aetch	cheat, teach	naptl	plant
racig	cigar	aspyn	pansy
tegas	gates, stage	metss	stems
ribet	tribe	alilc	lilac
lnein	linen	vreri	river
iplsl	pills, spill	nirsa	rains
mesoc	comes	ypppo	poppy
jonye	enjoy	charn	ranch
sracf	scarf		
armch	march, charm	*Eating-set Training*	
lsalh	shall, halls	plepa	apple
arypt	party	ackse	cakes
oautb	about	skanc	snack
tirew	write	upenr	prune
wlde	dwell	fiken	knife
admle	medal, lamed	teibs	bites
nuecl	uncle	nooni	onion
ngrbi	bring	sdoas	sodas
ownsk	knows	gsrau	sugar
woetl	towel	oabnc	bacon
euqen	queen	dnyca	candy
lichld	child	ocoac	cocoa
natri	train	vesno	ovens
eapll	lapel	acmer	cream
		reybr	berry
Nature-set Test		*Eating-set Test*	
sdyia	daisy	pyrus	syrup
donsp	ponds	dugfe	fudge
ekcer	creek	bracs	crabs
mawps	swamp	clnuh	lunch
sargs	grass	asadl	salad
nsswo	snows	turif	fruit
gofgy	foggy	rofsk	forks
ttyej	jetty	ragep	grape
rntho	thorn, north	yllej	jelly
eoanc	ocean, canoe	kinrd	drink

followed by two eating- and two nature-set anagrams in an ABBA order. Before each *S* in the weak habit-strength group was run, the eating and nature anagrams were selected at random from those used with the strong habit-strength group. The two groups were thus equated for the total number of training trials. This equating procedure was necessary, because preliminary experiments indicated that amount of practice was a significant variable influencing the subsequent solution of a particular class of anagrams. A test series of ten anagrams was given immediately following the conclusion of the training series. Half of the *S*s were tested with eating-set anagrams while half received nature-set anagrams.

A 2-min. time limit was enforced for each anagram during the training and test series. If *S* did not solve the anagram within 2 min. or gave an incorrect solution, a correct solution was given by *E*. The experiment was conducted in a small room containing only a table and chairs for *E* and *S*. The *E* placed each anagram card on the table in front of *S* and removed it from view before the next anagram was presented. The *S*'s solution time was determined with the aid of a stop watch.

The following instructions were given to *S*s of both groups: "This is an experiment on thinking. You are to solve the five-letter anagrams that I will present to you, as quickly as possible. A five-letter anagram is simply five scrambled letters. Your solution must use all the letters, that is, it must be a real five-letter word. It must not be a proper noun or slang. As soon as you have the answer, call it out to me and I will give you another to solve, but be sure your word fits the five letters before you name it. If you can't find a solution within 2 min. we will go on to another. Here is an example: . . . Now solve this one . . . and now this one. . . . Are there any questions? Do not ask any questions once the experiment has begun. Remember, call out your answer as soon as you find one."

Results

Table 2 shows the mean number of errors for the weak (W) and strong (S) habit-strength groups for Trials 1–5 and 6–10. The significance of the difference between Trials 1–5 and 6–10 in Cond. W was determined by the *t* test for related measures. A *t* of 2.66 was obtained. For 19 *df* it is significant at the .02 level, indicating that some learning had occurred during the test series. A similar test for Cond. S was unnecessary, since it showed a slight increase in mean errors during Trials 6–10. The *t* test of the difference between the differences of the two groups was 2.38 ($p = .02$, for 38 *df*). A nonparametric *u* test (7) indicated that the distributions of errors committed by the W and S groups on Trials 1–5 were not significantly different. A *u* of 1.83 was obtained for Trials 6–10 which is significant at the .03 level of confidence (one-tailed test). No significant differences were obtained between nature-set anagrams and eating-set anagrams for the test series.

TABLE 2

Mean Errors of W and S Groups in Trials 1–5 and 6–10 of Test Series

Trials	Group W		Group S	
	Mean	*SD*	Mean	*SD*
1–5	1.35	1.10	1.05	.74
6–10	.50	.67	1.15	1.20

Following the conclusion of the test trials *E* questioned each *S* with respect to his knowledge of the purpose of the experiment, whether he noticed any similarity or relationship among the anagrams, and his method of attacking the problems. None of the *S*s reported that they were aware of the purpose or design of the experiment. The protocols indicated that 75% of the *S*s tested with eating-set anagrams verbalized this similarity among the anagrams. Eighty per cent of the *S*s tested with nature anagrams verbalized this relationship. Ten per cent of the *S*s indicated some recognition of both eating and nature anagrams. The latter were all in the strong habit-strength group. Seventy-five per cent of the *S*s recognizing some similarity among the anagrams were in the weak habit-strength group. Eighty per cent were in the strong habit-strength group. There were no significant dif-

ferences in errors between *Ss* verbalizing a similarity and those that did not.

Forty-eight per cent of the *Ss* reported that they attempted to solve the anagrams by some systematic method of letter arrangement such as vowels around consonants or some letter order. Forty-eight per cent of the *Ss* reported that for some of the anagrams they tried a method of rearrangement of letters while for other anagrams the solutions just suddenly appeared. The remaining *Ss* reported that the solutions "just came" to them. These different methods of attacking the problems were approximately equally distributed among the S and W subgroups. There was no significant difference in errors between the *Ss* reporting that they adopted one method or the other in attempting to solve the anagrams. Only one *S* reported that knowing the anagrams belonged to the same general class seemed to help him solve the anagrams.

Discussion

The results of the experiment reported support the implication concerning mental set drawn from Hull's formulation of the principle of habit strength. The authors' initial assumption was that if "mental set" designates behavior which can be explained in terms of the acquisition of differential amounts of habit strength by a class of stimuli for the elicitation of a particular class of responses, then deductions from the principle of habit strength should hold for the phenomena described as mental set. The implication tested in the present study follows from the negatively accelerated growth function of habit strength. It is to the effect that a difference in strength would develop more rapidly between two habit strengths of equal but low absolute strength as compared with two habit strengths of equal but high absolute strength. The absence of a significant difference in Trials 1–5 indicates that several trials were necessary before the difference between competing reaction potentials in the W group became significantly greater than the difference between competing reaction potentials in the S group.

As previously indicated, the problem of directed thinking is similar in some respects to that of the selection of the correct response in a trial and error learning situation. Descriptively, selective learning may occur when the ultimately correct response is not the dominant response in the initial response hierarchy of alternative responses elicitable in a given stimulus situation. Learning is said to occur when the correct response becomes dominant in the response hierarchy by acquiring the greatest amount of effective reaction potential.

Descriptively, another kind of learning, discrimination, is necessary when the correct cue or class of cues among a number of alternative stimuli is initially weaker than the alternative stimuli in its tendency to evoke a given response. Discrimination learning occurs when the correct cues become dominant by acquiring the greatest amount of effective reaction potential for elicitation of the given response (11).

A given nature anagram, for example, may serve initially as a stimulus for the arousal of nature words, nouns in general, verbs, particular letter orders, and so on. With repeated solutions of nature anagrams these other classes of responses are sometimes reinforced but are also nonreinforced or extinguished on other occasions. Only the class of nature responses is consistently reinforced. It therefore becomes dominant, presumably through mediated generalization (1), in the hierarchy of responses elicited by nature anagrams. However,

only nature anagrams are consistently reinforced when correct nature responses are elicited. No other stimulus complex is consistently reinforced following the elicitation of this class of response. Differential reinforcement of this class of anagrams will make it dominant in the hierarchy of stimuli eliciting nature responses.

Establishment of a mental set of the type studied here involves a change in both the initial response and stimulus hierarchies. It involves both trial and error and discrimination learning. The behavior described as set is thus the result of the acquisition of a differential amount of reaction potential by a particular class of stimuli (discrimination learning) for the elicitation of a particular class of responses (selective learning).

Thus all the variables determining reaction potential and the manner in which different reaction potentials interact may be determiners of the direction of thought. This implies that directed thinking is the consequence of associative factors, which may be called mental set, and nonassociative factors, which may be called direction or determining tendency. This study was primarily concerned with the former.

Finally, it may be noted that the acquisition of symbolic responses and their utilization in new stimulus situations may come about as the result of discriminatory and selective learning as outlined in the present paper. The principles employed in the account of mental-set phenomena may also account for concept formation (**10, 12**).

Summary

An experiment employing 40 *S*s was conducted in order to test an implication for the direction of human problem solving derived from Hull's assumption that habit strength is a negatively accelerated growth function. An inverse relationship between amount of training with incompatible anagrams and subsequent success in the solution of one of these classes of anagrams was predicted on the basis of this assumption. The experimental results were in accord with this prediction. A behavioral interpretation of the concept of mental set and directed thinking was outlined. The establishment of anagram sets was discussed in these terms.

(Received February 23, 1952)

References

1. Cofer, C. N., & Foley, J. P. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.
2. Gibson, J. J. A critical review of set. *Psychol. Bull.*, 1941, **38**, 781–817.
3. Hull, C. L. Simple trial-and-error learning: a study in psychological theory. *Psychol. Rev.*, 1930, **37**, 241–256.
4. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
5. Humphrey, G. The problem of the direction of thought. *Brit. J. Psychol.*, 1940, **30**, 183–196.
6. Maier, N. R. F. Reasoning in humans: III. The mechanisms of equivalent stimuli and reasoning. *J. exp. Psychol.*, 1945, **35**, 349–360.
7. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
8. Rees, H., & Israel, H. An investigation of the establishment and operation of mental sets. *Psychol. Monogr.*, 1935, **46**, No. 6 (Whole No. 210).
9. Siegel, P. S. The role of absolute initial response strength in simple trial-and-error learning. *J. exp. Psychol.*, 1945, **35**, 199–205.
10. Smoke, K. L. Concept formation. In P. L. Harriman (Ed.), *Encyclopedia of psychology*. New York: Philosophical Library, 1946.
11. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
12. Vinacke, W. E. The investigation of concept formation. *Psychol. Bull.*, 1951, **48**, 1–31.

EFFECT OF VARYING AMOUNTS OF REST ON CONVENTIONAL AND BILATERAL TRANSFER "REMINISCENCE" [1]

G. ROBERT GRICE

University of Illinois

AND

BRADLEY REYNOLDS

Perceptual and Motor Skills Research Laboratory, Human Resources Research Center

A phenomenon similar to reminiscence in pursuit rotor learning has been reported by Irion and Gustafson (8) when *Ss* performed with one hand prior to the rest interval, but changed to the other hand following the rest. A similar result has also been obtained by Ammons and Ammons (3). In the Irion and Gustafson experiment, two groups of *Ss* practiced for ten 25-sec. trials with the right hand followed by ten trials with the left. All trials were separated by 5 sec., except that for one of the groups a 5-min. rest was introduced before the shift in hands. The postrest performance with the left hand was significantly better for the rest group.

This finding is of considerable interest because it may cast some light on the nature and locus of reminiscence and other work and rest phenomena. For example, Hull's theory of reactive inhibition, which has been extended to motor skill learning by Kimble and others (9, 10, 11), assumes that the amount of inhibition and hence the amount of recovery possible is a function of the frequency with which the specific response involved has been made. This would seem to localize inhibition in the specific effector system, and the above finding would present an obvious difficulty. Of course, there is no assurance that this effect and ordinary reminiscence are basically the same phenomenon. One way of approaching this problem is to ascertain if ordinary and bilateral "reminiscence" are similar functions of the same variables. In the present experiment the two types of gain are studied as functions of the length of the rest interval.

The functional relation between reminiscence and length of rest in rotary pursuit learning has been studied by Melton (14), Bell (4), Ammons (2), Kimble and Horenstein (11) and Irion (7). The finding has been that the gain is a negatively accelerated increasing function of the duration of the rest period up to a maximum of 5 or 10 min. Kimble and Horenstein fitted a simple exponential function to their data. In the present paper, two parallel experiments are reported showing reminiscence and "bilateral reminiscence" as functions of time of rest from 10 sec. to 10 min. This is done in order that the two functions may be compared, and, incidentally, to

[1] This study was conducted in the Perceptual and Motor Skills Laboratory, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas. The present report was prepared primarily at the University of Illinois under the provisions of Air Force Contract 33(038)-25726 sponsored by the Human Resources Research Center. The opinions expressed are those of the authors and not necessarily those of the Air Force.

determine the law for ordinary reminiscence with added precision.

Method

Subjects.—The *S*s were 432 right-handed airmen basic trainees selected at random from those entering basic training at Lackland Air Force Base during a four-week period in August and September, 1951. To determine handedness, all *S*s who said they were right handed were screened by a simple test. They were handed a 3-ft. rod, and the hand with which they received it was recorded. They were then asked to show how they would use the rod as a bat, broom, rake, and shovel. Only those who were judged to perform all tasks in a right-handed manner were used as *S*s. These men were then assigned at random to the experimental groups.

Apparatus.—The apparatus consisted of 12 units of the standard Air Force Rotary Pursuit Test described by Melton (15). The units were arranged in two rows of six with *S*s facing each other, and all *S*s were run in groups of 12. Trials and rest intervals were timed automatically by electronic timing devices and time on the target was recorded for each trial in units of .001 min. on Standard Electric timers. A buzzer automatically sounded 2 sec. before the beginning of each trial.

Procedure.—Practice periods were 30 sec. in duration and were separated by 10-sec. rest periods. The variable rest periods were introduced after Trial 15. In the straight reminiscence experiment, *S*s employed the left hand before and after the rest (LL groups). In the bilateral transfer experiment, *S*s practiced with the right hand before the rest and the left following the rest (RL groups). Separate groups of 36 *S*s were run under each LL and RL condition. The rest intervals, which were the same in the two experiments, were 10 sec., 30 sec., 1 min., 3 min., 5 min., and 10 min. All groups received at least five trials following the rest, but the 10-sec., 5-min., and 10-min. groups received 15 trials in order that the postrest performance curves could be examined.

Standard Air Force rotary pursuit instructions were employed, with only the exceptions made necessary by the experimental conditions. At the end of the initial instructions, all *S*s were told, "You may be asked to change hands during the test. If you are, just change the stylus to the other hand and continue as before. You will begin with your right (or left) hand." Immediately at the end of Trial 15, the 10-sec. RL group was told, "Change the stylus to your left hand." The 10-sec. LL group continued without interruption. At the end of Trial 15, all other groups were instructed, "Put down the stylus and rest." Fifteen seconds before the time to resume practice they were told, "Pick up the stylus." Ten seconds before time to begin, the RL groups were told, "Change the stylus to your left hand." Thus, all of the RL groups had an equal amount of time to make the change in hands. Three *E*s were always present. One gave the instructions and proctored, while the other two operated the equipment and recorded the scores for each trial.

Results

If gain as a function of time was to be compared for the two situations in this study, it was necessary to devise some measure which would apply to both conditions. The usual method of indicating reminiscence by subtracting some measure of prerest performance from a comparable measure of postrest performance could not be applied. The RL groups' postrest performance with the left hand was below that of the prerest right-hand performance, while the usual gains were shown by the LL groups. The desired comparisons actually were between gains over *predicted* postrest massed-practice performance. The measures presented are of this type. For each

TABLE 1

Mean Values for Each Experimental Group

Groups	Total Trials 11–15	Predicted Total Trials 16–20	Obtained Total Trials 16–20	Mean Gain
LL				
10 sec.	545	675	675	0
30 sec.	499	628	666	38
1 min.	580	711	890	179
3 min.	534	664	979	315
5 min.	671	803	1211	408
10 min.	628	759	1191	432
RL				
10 sec.	866	373	373	0
30 sec.	815	349	405	56
1 min.	731	311	438	127
3 min.	838	360	498	138
5 min.	864	372	598	226
10 min.	873	376	745	369

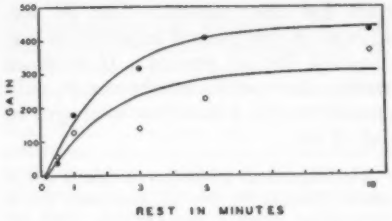

FIG. 1. Gain as a function of amount of rest. The filled circles represent the obtained values for the LL groups, and the unfilled circles represent the values for the RL groups.

of the two 10-sec. groups, correlation coefficients were computed between the total of the last five prerest trials and the total of the first five postrest trials. Linear regression equations were then obtained for each condition for predicting the postrest measure from the prerest measure. For the 10-sec. LL group r was .86, and for the 10-sec. RL group r was .77. The two regression equations were:

$$\text{LL: } y = 1.015x + 122$$
$$\text{RL: } y = .456x - 22$$

The gain scores presented are totals of the five postrest trials minus the predicted total for massed practice (i.e., 10-sec. rest). This measure is not only appropriate to the present study, but is probably preferable to the usual methods of presenting reminiscence data since it takes into consideration the correlation and variability.

The basic data in terms of mean values for each group are presented in Table 1. Gains as functions of rest are also presented graphically in Fig. 1. All data are in units of .001 min. on the target. The over-all trends for both conditions are highly significant as indicated by simple analyses of variance. Since the analyses were all computed from gains over predicted mean values, they actually are a form

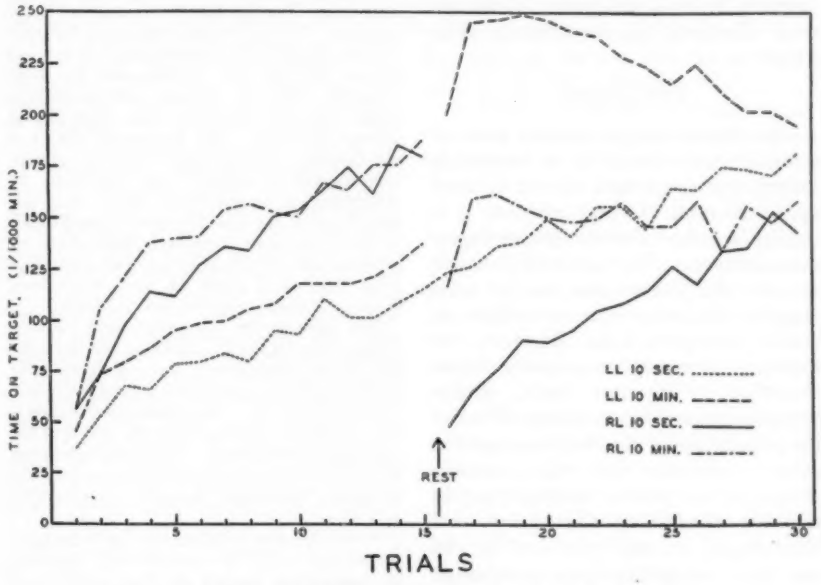

FIG. 2. Learning curves for the 10-sec. groups and the 10-min. groups

of analysis of covariance. All values were made positive for the analyses by the addition of a constant. The F value for the LL experiment is 8.88 and for the RL experiment it is 8.19. For the appropriate df these F's are both significant at the .001 level of confidence. In general, the gains for the LL groups are greater than those for the corresponding RL groups. An analysis of variance was applied to the difference between the two conditions yielding an F based on the between-rows variance with the within-groups variance as the error term. The F value of 5.80 is significant at the .05 level of confidence.

Learning curves for the two 10-sec. and the two 10-min. groups are presented in Fig. 2. Here the gains for each condition are illustrated by the postrest differences between the corresponding massed-practice and rest groups. It will be noted that for both conditions the beneficial effect of rest appears to be a temporary phenomenon, disappearing as practice continues.

Discussion

The results clearly indicate that bilateral "reminiscence" is an increasing function of the length of rest at least up to 10 min. In this respect, it is similar to the function for ordinary reminiscence. This fact still does not identify the phenomena, but at least suggests that they may be subject to similar interpretation. At least, the bilateral phenomenon requires interpretation in its own right. Other experiments will be necessary to study the possibility of similar functions for other variables. One thing accomplished by the present finding is to rule out the possibility that Irion and Gustafson's massed practice *S*s did not have adequate time to change hands, and merely got off to a bad start for that reason. The 10 sec. allowed in the present experiment was constant for all groups. If such an artifact had produced the result, gain should not be a function of length of rest at all.

Reminiscence as a function of time is such a regular function for the LL condition, that a curve-fitting attempt was justified. Hull (6) assumes that reactive inhibition is a simple exponential function of time.[2] While this theory, strictly speaking, is only a theory of classical and instrumental conditioning, it has been extended to motor skill learning by Kimble and his associates. Kimble and Horenstein (11) were able to fit such an exponential function to their data. Such a function has also been fitted to the present data, and, as may be seen by the upper fitted curve in Fig. 1, an excellent fit was obtained. The equation is:

$$Gain = 480\,(1 - 10^{-0.33r}) - 36$$

Interpreted in terms of the theory, the value of the asymptote (444) is determined by the total amount of inhibition accumulated during the prerest performance. The exponential decay function is then subtracted from the asymptote. The function indicates that *S*s recover from a constant fraction of the remaining inhibition per unit time. The fact that the G intercept is not zero but -36 may be explained by the fact that the rest period for the massed-practice group was 10 sec. and not zero. If the gains had been computed from a zero-rest massed-practice group, the curve could have been made to pass through the origin. The closeness of the fit of this function certainly supports such an assumption concerning recovery from inhibition. In attempting to extend Hull's theory into this area, however, one should remember that Ellson's (5) spontaneous recovery data, from which the postulate was derived, do not approach an asymptotic value until about an hour, while here, recovery is about complete in 5 or 10 min.

The gains for the RL groups were not only smaller than for the LL condition, but the function shows considerably less regularity. No simple negatively accelerated function will provide a very good fit. One possible explanation is that greater irregularity lies in the nature of the gain measures employed. The lower correlation between prerest and postrest performance for the RL group, means that the gains are computed

[2] Although stated differently, this is the same assumption made by Ammons (1) regarding his comparable concept of "temporary work decrement."

from less reliable predicted values. One plausible hypothesis is that the vertical differences between the two functions are genuine, but that the true functions are otherwise the same. That is, they are approaching different asymptotes, but each approaches its own limit at the same rate. To test this hypothesis, a lower asymptote was estimated for the RL function according to the method proposed by Lewis (12). This value was then substituted in the equation and the lower curve of Fig. 1 resulted. The new asymptote being 310, the equation became:

$$Gain = 346\,(1 - 10^{-0.22T}) - 36$$

A test of goodness of fit was then applied to test the hypothesis that the deviations of the obtained values from this function could be due to sampling error. The test, Case 7 suggested by Lindquist (13), involves a variance estimate based on "departure from hypothesis" and the within-groups variance as the error term. The test yielded an F of 1.60, which for 5 and 210 df is insignificant, and indicates that statistically the hypothesis is tenable.

Since the reactive inhibition theory assumes that inhibition results from the repetition of a specific response, it would at first appear that it cannot account for the gains obtained with the change in hands. However, it should be remembered that pursuit rotor performance is a very complex task, involving not only movements of the hand, arm, and shoulder muscles, but movements and postural adjustments of the entire body, as well as complex perceptual adjustments. Proponents of the theory could contend that recovery from inhibition localized throughout the body might well account for the gains. The difference between the asymptotic values of the two curves could be taken to represent the amount of inhibition localized in the initial hand and arm. A more simple and direct explanation of the difference between the two curves, however, could be simply that enough bilateral transfer did not occur to produce gains of the same magnitude, even though all inhibitory effects were dissipated. Reasoning of the above kind can be carried on with assurance only when something is known concerning the absolute level of learning. The present finding does at least indicate that there are temporary inhibitory factors which may involve gross bodily effects, perceptual factors, or perhaps central phenomena. Furthermore, these inhibitory effects dissipate as a function of time, perhaps in the same manner as the inhibitory effects which account for the conventional reminiscence phenomenon. Further experiments and theoretical analysis will be required to reveal the nature and locus of these inhibitory phenomena.

A further similarity between the two kinds of gains may be seen in the learning curves of the 10-sec. and 10-min. groups in Fig. 2. Here it may be seen that the superiority of both rest groups over their corresponding massed-practice groups diminishes, when practice is continued under the massed condition, and is entirely gone in from 12 to 15 trials. The difference between the two LL groups on Trial 30 is no larger than the initial sampling error between the groups. The curves for the 5-min. groups, which are not presented, show a similar picture with the exception of the fact that the gains are smaller. This may be taken to indicate that the inhibitory factors in both conditions reach a maximum value in a relatively short time. This is in accord with Irion's finding that reminiscence as a function of amount of prerest practice reaches a maximum value after about 10 to 20 25-sec. trials.

Summary

1. Two parallel experiments were conducted in which rest periods of 10 sec., 30 sec., and 1, 3, 5, and 10 min. were introduced after 15 30-sec. trials of rotary pursuit performance. In one experiment *S*s practiced with the left hand before and after the rest (LL condition), while in the other experiment they employed the right hand before the rest and the left hand afterwards (RL condition).

2. Gain over predicted massed-practice scores was an increasing function of amount of rest for both conditions. The LL groups showed larger gains.

3. For the LL conditions gain was a simple exponential function of length of rest up to about 10 min. The comparable function for the RL conditions showed less regularity, but the hypothesis is tenable that it is the same function approaching a lower asymptotic value.

4. In both conditions the superiority of the rest groups over the massed-practice groups disappeared with continued practice.

5. It is suggested that the gains in bilateral transfer indicate the presence of temporary inhibitory factors either associated with generalized postural

adjustment or of a perceptual or central character.

(Received February 25, 1952)

References

1. Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
2. Ammons, R. B. Acquisition of motor skill: II. Rotary pursuit performance before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.
3. Ammons, R. B., & Ammons, C. H. Bilateral transfer of rotary pursuit skill. *Amer. Psychologist*, 1951, **6**, 294. (Abstract)
4. Bell, H. M. Rest pauses in motor learning as related to Snoddy's hypothesis of mental growth. *Psychol. Monogr.*, 1942, **54**, No. 1 (Whole No. 243).
5. Ellson, D. G. Quantitative studies of the interaction of simple habits. I. Recovery from specific and generalized effects of extinction. *J. exp. Psychol.*, 1938, **23**, 339–358.
6. Hull, C. L. *Principles of behavior*. New York: D. Appleton Century, 1943.
7. Irion, A. L. Reminiscence in pursuit-rotor learning as a function of length of rest and of amount of pre-rest practice. *J. exp. Psychol.*, 1949, **39**, 492–499.
8. Irion, A. L., & Gustafson, L. M. "Reminiscence" in bilateral transfer. *J. exp. Psychol.*, 1952, **43**, 321–323.
9. Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
10. Kimble, G. A., & Bilodeau, E. A. Work and rest as variables in cyclical motor learning. *J. exp. Psychol.*, 1949, **39**, 150–157.
11. Kimble, G. A., & Horenstein, B. R. Reminiscence in motor learning as a function of length of interpolated rest. *J. exp. Psychol.*, 1948, **38**, 239–244.
12. Lewis, D. *Quantitative methods in psychology*. Iowa City: The Bookshop, 1948.
13. Lindquist, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65–78.
14. Melton, A. W. The effect of rest pauses on the acquisition of the pursuitmeter habit. *Psychol. Bull.*, 1941, **38**, 719. (Abstract)
15. Melton, A. W. (Ed.) *Apparatus tests*. AAF Aviation Psychology Program Research Reports, Report No. 4. Washington, D. C.: Government Printing Office, 1947.

ON THE LIMITS OF STEREOSCOPIC VISION [1]

KENNETH N. OGLE

Section of Biophysics and Biophysical Research,
Mayo Foundation, University of Minnesota,
and
Mayo Clinic

That stereoscopic perception of depth may exist even when the disparities between the images of the two eyes are large enough that those images are perceived double had been known before Helmholtz's time (**4**, v. 3, p. 430–431; **5**, p. 221). Yet much of the subsequent literature is not clear on this point. Hillebrand (**6**, p. 205) wrote that stereoscopic perception of depth occurs only so long as the disparate images are fused. Matsuda (**9**) presented data to show that under low illumination the limits of stereopsis agreed with the limits of fusion. Yet those familiar with the Hering "drop test" know that the falling bead can unequivocally be seen behind or in front of the fixation point, even when the images of the falling bead are seen double. Tschermak (**12**) took the position that fusion of the disparate images represents the most favorable condition for stereoscopic depth localization, but not the absolute condition.

Experiments by Tschermak and Hoefer (**13**), by Aall (**1**), and by Burian (**2**), to select a few examples, tended to show that even relative stereoscopic localization could be quite precise, even though the test objects were seen in double images. Helmholtz and others have pointed out, however, that if the disparity is too great, stereopsis will not exist.

Estimates of the limiting disparity were made by French (**3**), who used two pairs of vertical lines, the separation of the lines in each pair being different. Without using a stereoscope, but observing these lines in "free" stereoscopy, he found that, though two of the lines were perceived double, he had stereoscopic depth perception except when the ratio of the separations exceeded 1.6, for then depth perception ceased. Previously, Prandtl (**11**) in studying the Panum phenomenon, in which a stereoscopic depth is perceived when one eye sees a single vertical line and the other eye sees a pair of vertical lines, reported that when the separation of the two lines exceeded 1° of arc, the phenomenon no longer occurred. Eye movements were, of course, permitted in both these experiments. A test of stereoscopic vision was devised by Koch (**7**) during the last war, based on measuring the limiting disparity for stereopsis. He used simple red-green anaglyphs with targets which gave the familiar pyramid observed toward the apex end. Disparities of as much as 4° were occasionally reported. Scanning eye movements were also permitted in this test. The specifications of the targets he used were not given, so that the basic limiting capacity of the eyes for stereopsis cannot be computed. Such a pyramid is not a simple enough configuration with which to make such calculations. No systematic investigation for the limiting values of stereoscopic depth perception appears to have been made. Yet the existence of such limits is an im-

[1] Read before the Division of Experimental Psychology, at the meeting of the American Psychological Association, Chicago, Illinois, Sept. 4, 1951.

portant fact in our understanding of the binocular visual processes.

Experimental technics used to study stereopsis usually cannot avoid some empirical factors to depth perception. As the test object is moved behind or in front of the plane of fixation, its angular size changes, and its brightness and color change. Moreover, its images in the two eyes blur, which not only gives a distance clue but also acts as a stimulus for a change in accommodation. The ranges of disparities available in these apparatuses were also restricted by mechanical limitations. With the stereoscope unnatural conditions exist, including perhaps such inhibiting factors as texture of paper on which the figures are drawn.

Instrumentation

The apparatus used in the experiment to be reported here eliminated all these empirical factors, and permitted a measurement of the range of disparities within which stereoscopic depth could be perceived in the visual plane, for central and peripheral vision.

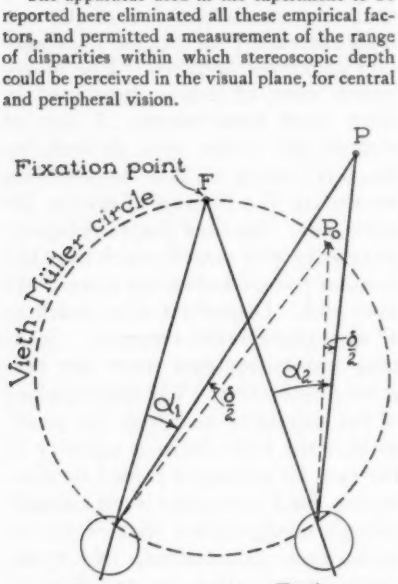

Fig. 1. Relationships between the spatial positions of points in space, the corresponding visual angles, and the geometrical disparity

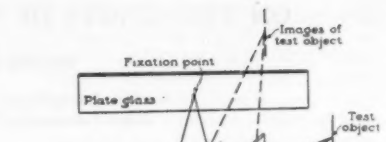

Fig. 2. Schematic sketch of the apparatus used to measure the limits of stereoscopic vision

Figure 1 shows the geometrical relationships between the disparity δ and the visual angles α_1 and α_2 which subtend the distance F–P to the left and right eyes respectively. The disparity $\delta = \alpha_2 - \alpha_1$ as measured here is the geometrical disparity, and as such is zero only when the point P lies at P_0 on the Vieth-Müller circle. The physiological disparity will differ slightly from δ, but the error in neglecting the difference even for a peripheral angle of 10° is only of the order of 15′ of arc (**10**). In view of the magnitudes of disparity measured with double images, this error will be neglected and the geometrical disparity used for the physiological.

The experimental apparatus used is illustrated schematically by the perspective sketch in Fig. 2. A fixation point consisting of a small illuminated spot, .25 mm. in diameter, is attached to a large sheet of thin plate glass, 50 cm. from the eyes. By reflection from special mirrors (50% transmission—50% reflection) at about an angle of 45° the eyes also see the images of a test object. The test object used for the data presented here was a thin, highly polished needle (about 3 cm. in length) suitably illuminated by a small projection lamp. The two mirrors can be made to rotate in *opposite* directions about axes *A*–*A*, by means of levers and a screw arrangement. A dial indicator shows the degree of turning of the two mirrors. A rotation of the mirrors introduces a disparity between the images of a point seen by the two eyes symmetrically about its position when lying on the Vieth-Müller circle (when the angle of disparity is zero). By turning the arm that supports the test object the disparity measurements can be made for a central and any peripheral visual angle. This instrument makes possible a large range of disparities of the images of the test object, and within this range the images do not change brightness or

angular size or become blurred. With a suitable headrest and chin cup, head movements are practically avoided. Horizontal slit apertures before the eyes and suitable distant backgrounds restrict the binocular visual field to the fixation point and the images of the test object. The experiment is conducted in a semidarkened room. Hence empirical clues to spatial localization are eliminated.

For a given experiment the observation of stereoscopic depth between the test object and the point of fixation corresponding to the angle of disparity could be made in three ways: (*a*) By means of a motor control, the mirrors could be rotated continuously and thus the disparity increased in one direction continuously. Or the change in disparity could also be introduced manually by a suitable knob (see Fig. 2), when the motor control is disengaged from the apparatus. (*b*) The images of the test object could be flashed (exposed) momentarily, by a tap key or telechron timing switch inserted in the circuit for illuminating the test object. (*c*) The image of the test object seen by one eye (its half-image) could be continuously seen while the half-image seen by the second eye could be exposed at the will of *S*, by means of a suitable sliding aperture mounted between the test object and the mirrors.

With steady fixation, the stereoscopic behavior of the images of the test object could thus be observed for given changes in the disparity, continuous exposure, momentary exposure and exposure of one of the half-images. Muscle imbalance between *S*'s eyes must be small in order that accurate fixation on so small a detail can be held. The principal difficulty is to hold fixation steady, while observing double images and stereoscopic depth. Some preliminary training is usually necessary.

Results

Subjective experience of stereoscopic depth.—Using the several procedures outlined in the preceding paragraphs, the experience of stereoscopic depth, observed as the disparity is increased in one direction, seems to fall into three ranges of disparity.

The test object is first seen at the same distance as is the fixation point. Then while an increasing uncrossed disparity is introduced, the image appears to move back, and *S* is aware of a strong stereoscopic sense of depth difference between fixation point and object—a sense of plasticity. The test object *appears* to move back proportionally to the increase in disparity. Then the images suddenly appear as two. At this point the disparity between the images has reached the outer boundary of the region of binocular single vision, and the disparity at this point corresponds to the horizontal extent of Panum's fusional area (**10**).

The stereoscopic sense of depth continues to increase, however, as the disparity increases, though the image of the test object is seen double. The separation of these images, of course, continually increases also. Then rather suddenly the strong sense of depth fades, and the depth interval no longer increases. Yet the two half-images still appear farther away than the fixation point. This disparity represents the limit of a patent stereoscopic experience.

With a continued increase in disparity, the separation of the double images also increases, and while no real sense of stereopsis exists, still the *two images unmistakenly appear farther than the fixation point.* Then rather abruptly even this impression ceases, and each of the two half-images appears indefinitely localized, and their separation merely increases. It may be questioned whether within this last range of disparities there is a true stereoscopic sense; yet there certainly is still a cerebral interaction between the images as seen by the two eyes, which gives rise to an impression of depth.

If the disparity is increased in the opposite direction, the stereoscopic perception of depth is one of nearer than the fixation point, but the same regions can be differentiated.

Measurement of the limits.—For the most part, the measurements of the disparities corresponding to the limits discussed in the preceding paragraphs

were obtained by the method of adjustment. In several instances the method of constant stimuli was also used, primarily to learn that such a procedure could be applied to the determination of these limiting disparities. There can be no doubt that these limiting disparities are not sharply defined, and under given conditions there would be a probability function of experiencing or not experiencing depth perception for a particular range of disparities.

To a certain extent the limiting disparities vary with the conditions of the experiment—general illumination, size of test object, fatigue of *S*—and probably somewhat with training. Failure to keep exact fixation is a small source of error, in that this would effectively increase the range of stereoscopic depth perception between limiting disparities. The amount of fusion detail in the fixation object as used here is less than that of the test object. Hence the disparity between the images in the test object seen by the two eyes would provide a strong stimulus for a fusional movement, and this would tend to weaken the hold of fixation on the dot. Momentary exposure of one or both of the half-images tends to eliminate this problem.

Two *S*s served as observers. One of the two was a trained observer of stereoscopic phenomena; the other was entirely lacking in such experience. The data were obtained in three ways as regards the exposing of the test object, namely:

1. The stereoscopic depth experience was observed for a continuous and regular change in disparity introduced by the electric motor driving arrangement. This movement was abruptly stopped the instant a given limit was judged to have been reached. A braking arrangement assures positive stopping of the movement.

2. The images of the test object were momentarily exposed by a tap key, following adjustment of the disparity either by motor control or by manual operation. The duration of the exposure was about a third of a second.

3. With one of the half-images continuously seen, the second half-image was momentarily exposed by means of a manual adjustment of the sliding aperture, again as the disparity was changed manually or by motor control. Sometimes a suppression of the unoccluded image occurred when the occluded second half-image was exposed. This of course interferes with the judgment of stereoscopic depth, and this phenomenon seems to occur more often when the disparity is near the limits of a patent stereoscopic sense of depth. For disparities not too large, it was a striking phenomenon to observe that the continuously seen half-image would be indefinitely localized, or localized at the will of *S*, yet when the second half-image was exposed, both would appear to jump back (or forward) in space stereoscopically under the stimulus of the disparity.

The data for the limiting disparities corresponding to the range of patent stereoscopic depth and those corresponding to the additional range where each of the double images is perceived qualitatively in depth are illustrated graphically in Fig. 3, for central vision and for peripheral angles out to about 7° of arc. The points plotted on the graph represent the average of the "far" (uncrossed) and the "near" (crossed) limiting disparities. This set of data was obtained by the untrained *S*. The magnitudes of the limiting disparities for the trained *S* were on the whole somewhat larger than for the untrained *S*.

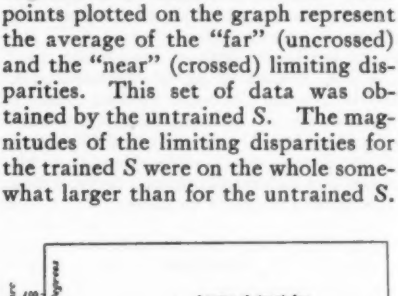

FIG. 3. Graphic representation of the data obtained for the limits of disparity corresponding to fusion, patent stereoscopic perception of depth and the qualitative appearance of depth of double images, showing also the changes with peripheral angle (data of the untrained *S*)

The disparities corresponding to the limits of fusion (and therefore the horizontal dimension of Panum's areas) are also shown, these being represented by the set of data lowest on the graph. The limiting disparities which define the range within which there is a patent stereoscopic perception of depth are shown by the two sets of data next higher on the graph, the lower set being the data obtained by momentarily exposing one of the half-images, the upper set being those obtained by momentary exposure of both half-images simultaneously. The set of data highest on the graph is for those limiting disparities which define the range where a qualitative depth of both of the double images is appreciated. These last limiting disparities varied less with the method of exposing the half-images. These limits indicate the largest disparities within which an interaction occurs between the neurological and cerebral processes of the two eyes.

The limiting disparities for patent stereopsis and those for a qualitative depth for double images differ somewhat with the time of exposure of the disparate images. Certainly with continued observation of the double images, the stereoscopic perception may fade and the images may be indefinitely localized—a fact mentioned in the literature. It is believed that the data shown here are quite representative of the magnitude of the limiting disparities.

A study of these data shows that even at the maculae of the two eyes there is a range of disparities outside of which there is no patent stereoscopic perception of depth, a range which is only slightly greater than the size of Panum's areas for the same region of the retinas. The measurements here of Panum's fusional area are of the same magnitude as those values previously found (**10**). A qualitative appearance of depth of the two double images increases the range by about 5′ of arc. Since the experimental arrangement of the apparatus causes these disparities to increase symmetrically with respect to the fixation point, the cortical representation of the two half-images would be, according to presently accepted theories, on the opposite lobes of the occipital cortex.

The range of disparities for patent stereopsis increases quite rapidly with the peripheral angle, and the range for a qualitative appreciation of depth of the double images increases even more rapidly.

The range of disparities within which there is a patent sense of stereoscopic depth is larger when the half-images are exposed simultaneously than when one of the half-images is seen continuously and the other is exposed momentarily. In the case of the simultaneous exposure of the two images, a subjective identification of the two images as belonging to the same external object might be the cause of this difference, or it might be a "retinal saturation" (**8**) from the continuously seen half-image which inhibits the binocular interaction with the second half-image when it is exposed. The suppression of the continuously seen half-image occurs much more frequently in the exposure of one of the half-images. To prevent after-images and retinal fatigue, *S* was asked to look away from the fixation point between observations.

The spatial regions corresponding to the ranges of disparities of Fig. 3 can be obtained by computation.[2] These

[2] Very approximately the range of depth (Δb) between the fixation point and the boundary given by a limiting disparity (δ) can be calculated from the relationship $\Delta b = b^2\delta/2a$ in which b is the observation distance and $2a$ is the inter-

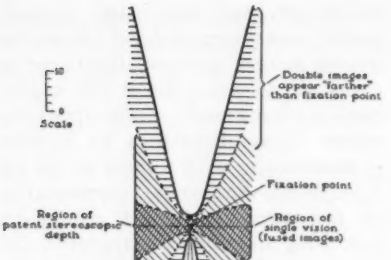

FIG. 4. The spatial regions corresponding to the several limits of stereoscopic depth perception calculated from the data of the trained *S*

regions calculated from the data for the trained *S* are shown in Fig. 4. An average of the disparities for the limiting values for the range of patent stereopsis was used. The region corresponding to binocular single vision (Panum's areas), that to patent stereoscopic depth and finally that to a qualitative depth perception with double images can be easily identified. Within the wedge-shaped (uncross-hatched) regions symmetrical with the median plane, in front and behind the fixation point, no stereoscopic depth perception would exist.

If one assumes that these limiting disparities are constant for all observation distances, then as the distance of the fixation point from *S* increases, the size of these spatial areas will also increase greatly—nearly with the square of the observation distance. If the eyes are fixed on a very distant object, the range of patent stereoscopic depth perception at a peripheral angle of about 6° will be from infinity to a distance of about 10 ft. from *S*. The effective range of patent stereoscopic vision will be increased if eye movements are permitted and an intermediate fixation point is selected.

The data presented here do not imply that there is a precise stereoscopic spatial localization from double images (though certain experiments cited in the literature would indicate this), but they do show that there is a stereoscopic perception of depth over a wider region than previously thought possible. Within one's ordinary surroundings, of course, the empirical factors of distance localization play a marked role independent of stereoscopic depth localization.

If stereopsis were entirely a psychological phenomenon, no limiting ranges of disparities as found would be anticipated. The fact that these finite ranges exist certainly is evidence for the existence of neuro-anatomical limitations, and probably a physiological basis for the stereoscopic perception of depth. The extent to which the neural paths that originate in the retinas are multiplied and, more pertinently, the extent to which these fibers overlap in the terminal regions of the occipital cortex (or some other area in the brain) would provide a basis for the neuro-anatomical limitations. Because at the macular regions of the two eyes a limiting disparity for stereopsis of the order of 10 to 20′ of arc is found, the neuro-anatomical overlapping probably cannot be, at least wholly, in the occipital areas of the cortex, because the half-images would be represented on opposite lobes of the occipital cortex.

Any generalized theory of space perception based on binocular vision must take into account the existence of these limiting disparities and their

pupillary separation. Specifically in the experiment here $\Delta b = .12\,\delta$ cm., when δ is expressed in minutes of arc.

corresponding region in space, outside of which there is no stereoscopic depth perception.

Summary

The fact that stereoscopic perception of depth is possible from disparate images which are seen double has been known since the middle of the last century. No systematic investigation of the limiting disparities within which such stereopsis exists, however, has been reported. In this paper an apparatus is described which permits such limiting disparities to be found without its introducing at the same time any of the usual empirical factors to spatial localization. Data on two *Ss* show that there are two ranges of depth with double images, the one in which there is a strong patent experience of stereoscopic depth, the second one in which there is a qualitative sense that both double images are seen farther or nearer than the fixation point, depending on the direction of the disparity. Measurements are made by the method of adjustment for the limiting disparities corresponding to these ranges. For a patent stereopsis these disparities range from about 10 minutes of arc at the maculas to 70 minutes of arc at a peripheral angle of 6 arc degrees. The qualitative sense of depth of the two double images ceased at disparities of about 15 minutes of arc at the maculas to nearly 2 arc degrees at a peripheral angle of 6 arc degrees. The extent of these regions has never been appreciated before. The existence of these limiting disparities is taken as evidence of neuro-anatomical limitations to be accounted for on the basis of the multiplication and overlapping of neural fibers in the terminal region of the occipital cortex or some other area of the brain.

These limiting disparities and the corresponding region in space outside of which no stereoscopic vision exists must be taken into account in any general theory of space perception based on binocular vision.

(Received March 4, 1952)

References

1. Aall, A. Über den Massstab beim Tiefensehen in Doppelbildern. *Ztschr. Psychol. Physiol. Sinnesorg.*, 1908, **49**, 108–127; 161–205.
2. Burian, H. Studien über zweiäugiges Tiefensehen bei örtlicher Abblendung. *Arch. Ophth.*, 1936, **136**, 172–214.
3. French, J. W. Stereoscopy re-stated. *Trans. Optic. Soc., London*, 1923, **24**, 226–256.
4. Helmholtz, H. von. *Treatise on physiological optics.* (Translated by J. P. C. Southall.) 1925.
5. Hering, E. *Spatial sense and movements of the eye.* (Translated by Carl A. Radde) 1942.
6. Hillebrand, F. *Lehre von den Gesichtsempfindungen auf Grund hinterlassener Aufzeichnungen.* 1929.
7. Koch, E. Ein neures Raumseh-prüfgerät. *Luftfahrtmedizin*, 1941, **5**, 317–321.
8. Köhler, W., & Wallach, H. Figural after-effects; an investigation of visual processes. *Proc. Amer. phil. Soc.*, 1944, **88**, 269–357.
9. Matsuda, A. Untersuchungen zur optischen Raumwahrnehmung. *Ztschr. Psychol. Physiol. Sinnesorg*, 1930, **61**, 225–246.
10. Ogle, K. N. *Researchers in binocular vision.* 1950, p. 345.
11. Prandtl, A. Die spezifische Tiefenauffassung des Einzelauges und das Tiefensehen mit zwei Augen. *Fortschr. Psychol.*, 1917, **4**, 257–326.
12. Tschermak, A. Optischer Raumsinn. *Handb. normal. path. Physiol.*, 1930, **12**, 834–1000.
13. Tschermak, A., & Hoefer, P. Ueber binoculare Tiefenwahrnehmung auf Grund von Doppelbildern. *Arch. ges. Physiol.*, 1903, **98**, 299–321.

INFORMATION THEORY AND IMMEDIATE RECALL

MURRAY ABORN AND HERBERT RUBENSTEIN

Michigan State College

As ordinarily used, the word *information* involves the idea of a message adding something to the addressee's knowledge. In communication theory, the concept of information is more restricted: only when the addressee learns something from a particular segment of message which could not be learned from the context in which the segment occurred, is the segment said to contain information.[1] The term information will be used only in this restricted sense throughout this study (**2**; **3**, p. 74–112).

The relation between context and information may be clarified by the following example. Suppose we heard the utterance "The door was hermetically ——." There would be no difficulty in getting the import of the utterance. We would automatically supply the word "sealed" since the occurrence of this word could be predicted with certainty from the portion of the utterance which we heard, i.e., from its context. "Sealed" is the only word which can occur in the context "The door was hermetically——." Since the utterance "The door was hermetically ——" functions just as well as "The door was hermetically sealed," we can say that "sealed" adds no information in this context.

Suppose we heard the fragment "The letter was ——." We could not be certain what the next word was. It could be "sealed," but it could also be "written," "mailed," "here," "on," etc. Since the occurrence of "sealed" here cannot be predicted with certainty from the context, it is clear that its presence tells us something we would not know if it were absent. Its presence, therefore, conveys information.

The information contained in a segment, then, is related to the probability of predicting its occurrence from what has gone before. The probability of predicting the occurrence of a certain word in a given context is, in turn, partially dependent on the number of words which may occur in that particular context. Mathematically stated, *the amount of information contained in any given segment is equal to the logarithm to the base 2 of the number of alternatives which can occur in the given context.* Thus, given a segment occurring in a context in which 15 other segments can replace the given segment (so that counting the segment which does occur, there are 16 alternatives for that particular context), we would say that the information value of the given segment in that context is $\log_2 16$ or 4 bits. The *bit* (a telescoping of *binary digit*) is the unit of information.

This is not the whole story. Given the context "I saw a very interesting ——," the statement would contain more information if the blank were filled by "antimacassar" than if it were filled by "play" since the occurrence of "antimacassar" is far less probable in this context than "play" is (and consequently far less predictable). If the alternatives in a given context differ in probability of occurrence, these differences must be taken into account in computing the amount of information that each of them contains. Such differences are always found in social languages. Since no data are available on probability of occurrence for any language, a linguistic information study at present can be carried out accurately only in acultural sequences such as nonsense material. The mathematical formula given above, which assumes the equal probability of all alternatives, is fully applicable to such material.

[1] Only the portion of the message preceding the segment in question is considered relevant in computing the amount of information contained in the segment.

To sum up the concept of information, we can say: (*a*) The amount of information conveyed by a particular segment can be calculated only in reference to a particular context. (*b*) It is a logarithmic function of the number of alternatives which may occur in that context. The greater the number of alternatives, the greater the amount of information conveyed by the given segment. Thus, randomness, i.e., freedom of occurrence of the segments composing our universe, results in high information values. Information and organization, i.e., limitation on the occurrence of the segments composing our universe, stand in an inverse relationship. (*c*) Differences in probability of occurrence, where they exist, must be taken into account in the calculation of the amount of information. Differences of this sort are found in all systems of social communication. Information studies in meaningful language aiming beyond rough approximation must await further data on these probabilities.

Statement of the Problem

The present study seeks to apply the concept of information to the problem of learning as measured by immediate recall. Specifically, information theory was used to furnish rules under which test passages of varying degrees of contextual organization were constructed. By the use of these rules, the information conveyed by the passage could be specified exactly.

A different, though related, gauge of organization was used by Miller and Selfridge (1), who graded the organization of their test passages according to the approximation to English. While this method enabled them to deal with meaningful material, degree of approximation to English, especially as worked out in their study, is not nearly so accurate a measure of contextual limitation as is provided by information theory.

The organization of our material is achieved by progressively limiting the context in which the various nonsense syllables can appear. The limitations applied in each passage constitute the organization pattern of the passage. The degree of organization is measured by the average rate of information per syllable. *The average rate of information is computed by dividing the total amount of information contained in the given passage by the number of syllables in the passage.* If the organization patterns are utilized by the *S*s, they should be able to recall a greater number of syllables in the more highly organized passages. Since, however, the rate of information per syllable is in inverse proportion to the degree of organization, the information score should be the same for all passages. More formally stated then, the hypotheses of this study are: (*a*) The amount of learning in terms of syllables recalled is greater as the organization of the passage is greater, i.e., as the average rate of information is smaller. (*b*) The amount of learning in terms of the information score, computed as the product of the number of syllables recalled and the average rate of information, is constant for all passages.

Method

Structure of the Nonsense Syllables

A vocabulary of 16 readily pronounceable nonsense syllables was constructed. These syllables were:

Class 1	*Class 2*	*Class 3*	*Class 4*
ZAP	YAV	KER	EMP
ZAS	YAD	KEZ	ERT
ZOP	YOM	KIF	IPS
ZOT	YOG	KIB	INF

In order to permit the setting up of simple rules of distribution, the nonsense syllables were

divided into four classes, each containing four syllables. To facilitate their learning, the syllables were constructed so that the membership of each class was characterized by some formal feature: Class 1 syllables all began with Z, Class 2 syllables with Y, Class 3 with K, and Class 4 with E or I. The other formal features may be seen by inspection.

Structure of the Passages

Six passages were constructed, all about of the same length, 30–32 syllables.[2] All syllables were employed with approximately the same frequency in the passages so that the differences in probability of occurrence played no role in the calculation of the amount of information. The passages differed from each other in the degree of organization, i.e., more limitations were imposed on the distribution of the syllables as one went from Passage 1 to Passage 6.

In Passage 1 the syllables were presented randomly. This meant that at any point in the passage, *S* had to choose his syllable from 16 alternatives. The rate of information per syllable was therefore 4 bits.

In Passage 2 a low degree of organization was introduced: The syllables were divided by commas into groups of two. Within each group, the second member belonged to the next higher class than the first. Thus, if the first syllable of the group belonged to Class 1, the second belonged to Class 2. If the first belonged to Class 2, the second belonged to Class 3, and so on. The progression was in the form of a circle so that when the first syllable belonged to Class 4, the second syllable belonged to Class 1. In selecting the first syllable of any set, *S* again chose from 16 alternatives (4 bits). In deciding on the second number of the set, however, he had to choose from only four alternatives (2 bits) since he could infer the class to which this syllable belonged from his knowledge of the class of the first syllable. The average rate of information in this passage was, therefore, 3 bits per syllable.

The syllables were again marked off into groups of two in Passage 3. There were only two kinds of groups. One consisted of a Class 1 syllable followed by a Class 2 syllable while the other consisted of a Class 3 syllable followed by a Class 4 syllable. The kinds of groups occurred in random order. The number of alternatives, then, from which *S* chose the first member was eight (3 bits), i.e., he chose from the membership of two classes, 1 and 3. The number of alternatives involved in his choice of the second member was only four (2 bits) since the class of this member is inferable from the class of the first member. Thus, the average rate of information was 2.5 bits per syllable.

In Passages 4, 5, and 6, the groups were four syllables long. Each group of four syllables was marked off by commas. Within each group, the first syllable belonged to Class 1, the second to Class 2, and so on. This was the total organization in Passage 4. At any point in Passage 4, then, since it was possible to infer the class of the sought-after syllable from the class of the preceding syllable, *S* chose from four alternatives. The rate of information in this passage, then, was 2 bits per syllable.

The additional restriction in Passage 5 was that adjoining syllables should not contain the same vowel-letter. This restricted the choice in Classes 2 and 4 since these classes contained syllables with the same vowel-letters as 1 and 3 respectively. The choice in Classes 1 and 3, however, remained unrestricted since none of the syllables in these classes contain the same vowel-letters as the classes they followed (Classes 4 and 2 respectively). Thus, *S* was faced with four alternatives (2 bits) in selecting a syllable which he knew to belong to Class 1 or to Class 3. In selecting a syllable which he knew to belong to Class 2 or 4, on the other hand, he had to choose from only two alternatives (1 bit) since two alternatives were ruled out as containing the same vowel-letter as the preceding syllable. The average rate of information in this passage was 1.5 bits per syllable.

Passage 6 had the same organization as Passage 4 plus the following restrictions: The vowel-letters in the odd-numbered groups were AAEE, e.g., ZAP YAV KEZ EMP, while the vowel-letters in the even numbered groups were OOII, e.g., ZOP YOG KIF INF. Selection at any point, consequently, was limited to two alternatives. The rate of information in this passage was 1 bit per syllable.

Correctness of Recall

Two criteria had to be satisfied for a syllable to be considered correctly recalled: (*a*) unambiguous reproduction of the form of the syllable, and (*b*) accurate reproduction of the position of the syllable.

Since the information calculations were only on the syllable level and not on the letter level, i.e., there was no concern with, say, the number of alternatives from which the *S* selected P after the ZA, there was an attempt to facilitate the correct reproduction of the syllable forms as much as possible. Thus, each *S* was required to memorize the inventory of syllables before the study was started. Further, this inventory was

[2] Passages 1, 2, and 3 each contained 30 syllables while 4, 5, and 6 each contained an additional two syllables in order that the passages might not end with an incompleted pattern.

printed on the top of each answer sheet for *S* to refer to at will. Errors in reproduction of form were, consequently, very few. If the reproduction was unambiguous, though inaccurate, e.g., AMP for EMP, credit was given provided, of course, the second criterion of correctness was also satisfied.

Position could be acceptably reproduced in either of two ways: (*a*) By reproducing the syllable in the same numerical position as it was in the passage. Numerical position, which can be considered an aspect of context, was completely adequate for the calculation of the information conveyed by the given syllable. Accordingly, if *S* reproduced ZAP correctly as the third syllable of the passage, it was considered correct even if the first and second syllables were incorrectly reproduced. (*b*) In a sequence of syllables which correctly reproduced the relative order of the passage syllables, the position of all syllables other than the first was considered adequately represented even though this position was incorrect numerically. The first syllable of such sequences had to be considered incorrect since it was improperly selected from the viewpoint of its preceding environment. Since some control over the order of the group as a whole was considered necessary, the degree of deviation from the numerical position was limited to a 4-syllable disparity. (In the last three passages, this was tantamount to saying that the reproduced sequence could be no more than one group away from where it was in the passage.) Thus, if syllables occupying positions 13, 14, 15, 16 were recalled as occupying positions 21, 22, 23, 24, the reproduction would be considered totally incorrect. If, however, they were recalled as occupying positions 17, 18, 19, 20, the syllables in positions 18, 19, 20 would be considered correctly reproduced.

Procedure

Group I.—The *S*s were all undergraduate students attending a course in psychology. As a group, they were quite homogeneous with respect to age, level of intelligence, and educational background. There were about twice as many males as females. The *S*s were allowed about three days in which to memorize the nonsense syllables and the characteristics of the classes into which they were divided. At the end of that time, they were given a criterion test consisting of items ranging from simple class identifications to the construction of syllable sequences based upon complex class combinations. Only 26 of the initially selected 40 *S*s were subsequently judged sufficiently proficient in the manipulation of the syllables to participate in the experiment. These 26 *S*s compose Group I.

In order to minimize the possibility of negative transfer between the organization patterns, Group I was administered one passage per class meeting, i.e., three to four days apart. To nullify the effect of practice, the order of the passages was randomized.

Each *S* was given a sheet containing the organization pattern of the passage on which the *S*s were to be tested and allowed a 10-min. study period to fix them in his mind. At the end of the study period, the first sheet was removed from sight; each *S* received a sheet containing the actual passage to be memorized, and allowed 3 min. to learn as much of it as he could. When 3 min. had elapsed, the passage was removed from sight and *S*s recorded as much of the passage as they could recall. This procedure was followed for all six passages.

Group II.—The data obtained with Group I suggested certain modifications in design which might serve to reduce variability. The change consisted of having *S*s take all passages in one sitting. To minimize fatigue, only four of the passages were employed (1, 2, 4, and 5). In other respects, the experimental procedure was identical to that followed with Group I.

In a new group of 20 students who were given the syllables to study, only seven were found sufficiently proficient to participate in the experiment. These seven *S*s compose Group II.

Results and Discussion

The data obtained for the performance of Group I are given as part of Fig. 1. The means for the number of

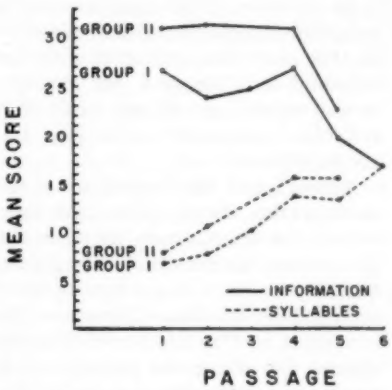

Fig. 1. Mean scores for number of syllables correct and information on the organization pattern passages

syllables correct are presented for each passage. They represent the raw scores from which the information scores were computed. As hypothesized, the means for number of syllables correct show a progressive increase in magnitude from Passage 1 (no organization) through Passage 6 (high organization), with the exception of Passage 5. Since *Ss* took the passages in random order, the increase could not be the effect of practice.

Were the increases proportionate to the average rate of information contained in each passage, the mean information scores would remain constant from Passages 1 through 6. As indicated in Fig. 1, the mean information scores fluctuate from passage to passage; the variation among means for Passages 1 through 4 being relatively slight (within a range of 3 score units), and that for Passages 5 and 6 being relatively large. In the case of Passage 5, the mean number of syllables correctly recalled shows no increase over the mean number recalled in Passage 4 and the mean information score drops considerably. In the case of Passage 6, the mean information score shows a further decrease despite the large increase in the mean number of syllables correctly recalled. An information score commensurate with that achieved for Passage 1, for example, would require an average recall of 26 syllables compared with the 16.5 accomplished.

According to the hypotheses of this investigation, the organizational structure of the six passages would enable *Ss* to memorize the syllables in greater quantity as the average rate of information becomes less. Moreover, the increment in the number of syllables learned for any given passage would result in an information score for that passage equal to the information score achieved for the previous passages.

TABLE 1

ANALYSIS OF VARIANCE OF INFORMATION SCORES OF GROUP I FOR PASSAGES 1, 2, 3, AND 4

Source of Variation	*df*	Mean Square	*F*
Passages	3	53.33	—
Subjects	25	113.28	1.34
Interaction	75	84.48	

The data in Fig. 1 testify that this did not occur. The means for Passages 5 and 6 seem particularly divergent, and the use of independent *t* tests for related measures showed that both are significantly different from Passages 1, 2, 3, and 4. In the case of the first four passages, however, the differences among means seem within the range of chance fluctuation. The notion that the data for these passages are statistically homogeneous is, in fact, a tenable one. When the information scores for Group I are subjected to an analysis of variance, the results (Table 1) indicate that neither passages nor subjects contribute significant variance when tested against the interaction variance.

The available evidence shows, then, that the amount of information contained in the material recalled was constant in the less organized passages. In the more highly organized passages, where the average rate of information fell below a critical value, the amount of information contained in the material recalled was reduced.

This is, in essence, the same finding as revealed by Miller and Selfridge (1), who observed that raising the orders of approximation affected the percentage of recalled material only up to a certain point. For example, in the ten-word passage the percentage recalled does not increase appreciably above the second order of approximation while in the 50-word passage the

percentage recalled does not increase appreciably above the fifth order of approximation.

The results of Group II are given as part of Fig. 1. The curves for Group II reflect a higher level of performance but are otherwise similar to those for Group I in most respects. There is the same "leveling off" of the curve for syllables correct after Passage 4 and the same consequent drop in the information curve. The curve for syllables correct shows a more consistent acceleration from Passages 1 through 4, however, and the information curve is correspondingly straighter (the mean information scores fall within a range of ½ of 1 score unit).

The results of an analysis of variance (Table 2) bear out the impression gained by an inspection of Fig. 1. The mean information scores for Group II show no significant variation for Passages 1 through 4. On the other hand, much less information was contained in the material recalled from Passage 5. The difference between Passage 5 and the less organized passages was found to be even more significant with this group than it was with Group I. Individual differences among the *Ss* reach a barely significant level in this group. This appears to be a consequence of a somewhat greater range in individual ability in the group of *Ss* chosen.

TABLE 2

ANALYSIS OF VARIANCE OF INFORMATION SCORES OF GROUP II FOR PASSAGES 1, 2, AND 4

Source of Variation	*df*	Mean Square	*F*
Passages	2	.50	—
Subjects	6	237.00	4.01*
Interaction	12	59.17	

* Significant at the .05 level.

In general, the results with Group II confirm the findings with Group I rather well. The altered procedure was moderately successful in reducing variability and the same results emerge in spite of the small number of *Ss*.

SUMMARY AND CONCLUSIONS

The present paper reports an attempt to investigate the influence of the degree of organization upon the ability of *Ss* to recall lists of syllables immediately after learning. A group of *Ss* was made familiar with a small inventory of nonsense syllables and later learned various rules of organization, each set of rules progressively limiting the syllables which could appear. They then tried to memorize passages of between 30–32 syllables chosen in conformity with the rules learned previously. Organizational arrangements employed ranged from random order (high average rate of information) to high organization (low average rate of information).

1. More syllables are correctly recalled from a passage with a lower average rate of information than from a passage with a higher average rate of information.

2. Up to a certain degree of organization, between 2 and 1.5 bits per syllable, the amount of information learned by the *Ss* was constant.

3. Beyond a certain degree of organization, about 1.5 bits per syllable, the relationship between the quantity of material recalled by *Ss* and the average rate of information contained in that material tended to become disproportionate. Although an increasingly greater number of syllables was correctly recalled, the total amount of information contained in the syllables was smaller than in the less organized passages.

4. Up to a certain degree of organi-

zation, the quantity of material which *S* can recall can be predicted from a knowledge of the average rate of information of the given passage and *S*'s level of ability as measured in a passage having a different average rate of information.

(Received March 6, 1952)

References

1. Miller, G. A., & Selfridge, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–185.
2. Shannon, C., & Weaver, W. *The mathematical theory of communication.* Urbana: Univer. Illinois Press, 1949.
3. Wiener, N. *Cybernetics.* New York: Wiley, 1948.

INSTRUCTIONAL SETS AND HABIT INTERFERENCE

MONCRIEFF H. SMITH, JR.

University of Washington[1]

Our everyday behavior is composed in large part of simple, well-defined responses made to any one of a large class of stimuli. The stimuli may remain the same from day to day, but learning varies membership in the class of stimuli to which the response is made. From a practical point of view, therefore, it is of considerable importance to discover what happens to such behavior when the rules are changed, i.e., when the same response must be made to a new class of stimuli which may or may not overlap the old class. In this situation the external stimuli remain unchanged, and the only alteration is in the nonspecific stimuli introduced by verbal instructions or by differential reinforcement.

Most theoretical analyses of transfer or of discrimination reversal concern themselves with similarity relations between original and transfer test stimuli (**1, 2, 5, 6**) and do not really bear on the question asked here. However, insofar as they do cover the present situation they suggest that habit interference can be predicted from specific factors alone (degree of practice, intraserial stimulus similarity, etc.). Possible interactions between instructional sets and previous practice are rarely considered. The experiment reported here was designed to explore such interactions.

[1] This experiment was done at the Perceptual and Motor Skills Research Laboratory, USAF Air Training Command Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas. The opinions or conclusions contained in this report are those of the author. They are not to be construed as reflecting the views or indorsement of the Department of the Air Force.

Apparatus and Procedure

The stimuli used were the 26 letters of the alphabet, printed in block capitals and photographed on 35-mm. film, one letter to a frame. The negatives were then spliced to form a continuous film strip 260 frames long, and random in sequence except for the restriction that each letter appeared just ten times. Eight strips were made, varying only in sequence of letters. The letters were flashed on a beaded screen from a projection window behind the *S*s. An Eastman 35-mm. film strip projector was modified to present the letters automatically at the rate of one frame per second. Changing frames required .2 sec., so that each letter was visible for .8 sec. The screen was 11 ft. from the *S*s and the image of the letters (white on black) was 20 in. high on the screen.

The task was to respond by pushing a button whenever any one of eight predesignated letters appeared on the screen. The *S*s were tested in groups of four. They were seated at a long table which was divided by partitions into four compartments so that they were unable to see one another. The partitions did not prevent them from hearing one another, but interaction was minimized by assigning each of the four *S*s a different set of letters and by instructing them that this was the case. Each *S* had a response box before him, and recorded his responses directly on an Esterline Angus operations recorder. The *S*s were all airmen undergoing basic training at Lackland Air Force Base.

Nine groups of four *S*s each were given eight trials—four in the morning and four in the afternoon. Approximately 5-min. rest was given between trials within a set of four. A trial was one sequence of 260 letters. After a short practice series in which *S* responded once to each of four letters, each *S* was given a slip of paper with eight letters, and told to respond whenever any one of the letters appeared. Each *S* responded to the same set of eight letters for the first five trials.

On Trial 6 each *S* was given a new set of letters. For one *S* in each group the letters were the same that he had been responding to all along, but were arranged in a different order on the paper. The nine *S*s so tested will be referred to hereafter as the "same" group.

One other *S* in each group was given a new set of eight letters, none of which was the same

as his original set. These nine *S*s will be called the "different" group. The other two *S*s of each group were given new sets consisting of four letters from their original set of eight, and four new letters. The order of the old letters on the new slip of paper was scrambled. These 18 *S*s will be called the "half" group.

For Trials 7 and 8 all *S*s responded to the same letters they were given on Trial 6. Schematically, the design looked like this:

Group	*n*	*First 5 Trials*	*Last 3 Trials*
Same	9	A B C D E F G H	G C F A B H E D
Half	18	A B C D E F G H	I *B G* J K *E* L *F*
Different	9	A B C D E F G H	I J K L M N O P

The actual letters assigned varied from man to man. In all cases the selection of letters and their order (within the framework of the design) was determined by a table of random numbers.

The slips of paper containing the original letters were given to the *S*s about 1 min. before the beginning of the first trial. They could study the letters then, and also during the 5-min. rest between trials. The slips were available during trials, but the 1-sec. rate of presentation left little time for studying the letters. On the first transfer trial, *S*s again had about 1 min. to study the new letters before the trial began, with possible study during and between subsequent trials.

Results

The design allowed a comparison of errors from one condition to another, and between errors made to different types of stimuli within an experimental condition. In order to make such comparisons meaningful it was necessary to reduce all scores to a common unit. The measure used was the number of errors per *S* divided by the possible number of such errors. Table 1 gives the results, expressed both as mean number of errors per *S* and as percentages of the possible number of errors. The mean of all errors is also presented for each group. This is simply the sum of the means of the different types of errors.

On the first five trials for all three conditions there were 18 letters unassigned, each occurring ten times, allowing 180 opportunities per trial per *S* for errors of commission (responses to letters not currently assigned). The 9.3% errors of commission for the "half" group on Trial 1 is 16.8 divided by 180. For errors of omission, there were only 80 opportunities per *S* per trial. Thus the 14.6% error of the "half" group is 11.7 divided by 80.

On the last three trials the opportunities for error differed from one group to another. For the "same" group the percentages were calculated in the same fashion for all eight trials,

TABLE 1

Trial by Trial Average Errors per Subject and Per Cent Error (Number of Errors Divided by Opportunity of Occurrence)

Trial	Errors of Commission by Group									
	"Same"		"Different"				"Half"			
	Not Asgd.		Not Asgd.		Asgd. Tr. 1–5 Not 6–8		Not Asgd.		Asgd. Tr. 1–5 Not 6–8	
	Mean	%	Mean	%	Mean	%	Mean	%	Mean	%
1	14.7	8.2	12.9	7.2			16.8	9.3		
2	7.1	3.9	7.4	4.1			10.6	5.9		
3	3.2	1.8	2.8	1.6			6.5	3.6		
4	1.8	1.0	2.3	1.3			4.6	2.6		
5	1.8	1.0	2.4	1.3			4.4	2.4		
6	1.3	0.7	9.4	9.4	5.4	6.8	4.8	3.4	10.4	26.0
7	1.1	0.6	3.8	3.8	5.2	6.5	2.1	1.5	6.3	15.8
8	0.8	0.4	1.9	1.9	2.8	3.5	1.9	1.4	4.8	12.0

since these *S*s were assigned the same letters in a different arrangement.

For the "different" group, eight letters were assigned for the first five trials, and eight different letters for the last three, leaving only ten letters on which erroneous responses to unassigned letters could be made. Thus, 9.4% is 9.4 over 100. There were 80 opportunities per *S* per trial for making errors by responding to letters that had been assigned for the first five trials but were not assigned for the last three trials. There were also 80 opportunities per trial for making errors of omission.

The "half" group had only four new letters assigned on the last three trials, so that there were 14 unassigned letters that could be responded to. The 3.4% of such errors on Trial 6 is 4.8 divided by 140. For each of the other three categories of error there were 40 opportunities per trial.

Table 1 shows that the "same" group was influenced little, if any, by the change in arrangement of letters assigned. Errors of omission show a slight rise, but this is more than offset by a drop in the errors of commission, so that the average of all errors shows an almost unbroken fall.

The "half" and the "different" groups both showed a pronounced increase in errors from Trial 5 to Trial 6. In the case of the "different" group, the average of all errors on Trial 6 was identical to that on Trial 1—27.9 errors on both trials. For the "half" group, there were slightly fewer errors on Trial 6 than on Trial 1 (26.6 and 28.6, respectively), but the difference is far from significant ($t = .28$). From this analysis it would appear that the "different" and the "half" groups had responded in a very similar manner to the conditions imposed on them.

However, an analysis of the relative frequency of the different types of errors shows that this was not the case, and that the two sets of instructions not only had different effects in this experiment, but might be expected to produce different results in other situations.

The "different" group showed no evidence of negative transfer. The average of all errors they made on Trial 6 was the same as that on Trial 1. The breakdown into errors of commission and omission did not correspond quite so neatly, but in neither case did the difference between Trials 1 and 6 approach significance. It might be

TABLE 1 (*Continued*)

Errors of Omission by Group										Total Errors		
"Same"		"Different"				"Half"				"Same"	"Diff"	"Half"
Asgd. Tr. 1–8		Asgd. Tr. 1–5		Asgd. Tr. 6–8		Asgd. Tr. 1–8		Asgd. Tr. 6–8				
Mean	%	Mean	%	Mean	%	Mean	%	Mean	%	Mean	Mean	Mean
16.9	21.1	15.0	18.8			11.7	14.6			31.6	27.9	28.6
3.7	4.6	2.8	3.5			3.8	4.8			10.8	10.2	14.4
1.8	2.2	1.2	1.5			2.3	2.9			5.0	4.0	8.8
1.6	2.0	0.6	0.8			1.2	1.5			3.4	2.9	5.8
0.6	0.8	0.4	0.5			1.0	1.2			2.4	2.9	5.4
0.8	1.0			13.0	16.2	5.4	13.5	5.9	14.8	2.1	27.9	26.6
0.4	0.5			7.2	9.0	1.2	3.0	2.7	6.8	1.5	16.2	12.4
0.4	0.5			3.6	4.5	0.9	2.2	1.3	3.2	1.2	8.2	8.9

argued that this is not a valid comparison, and that five-trials practice on a similar task should have produced a gain in performance. The task is so simple that this is doubtful, and pilot studies involving responses to different sets of letters on each trial showed no practice effect.

Moreover, the internal evidence shows no sign of interference in this group. Fewer errors were made in response to letters that had previously been assigned than were made to letters never assigned (6.8% error on previously assigned letters, and 9.4% on letters never assigned). This difference between percentages is not significant ($t = .61$), but the direction of the difference indicates no evidence for the carry-over of old responses.

For the "half" group there was evidence of negative transfer. It does not appear in terms of an increase in the total number of errors from Trial 1 to Trial 6. There was actually a decrease in total errors from 28.6 per *S* on Trial 1 to 26.6 on Trial 6. This decrease is more than accounted for by the reduction in responses to letters that were never assigned. There was an average of 9.3% of such errors on Trial 1 and 3.4% on Trial 6. This difference is highly significant ($t = 4.53$). Errors of omission were not significantly different on Trials 1 and 6.

The negative transfer appears in the carry-over of previously practiced responses. There was an average of 10.4 such responses per *S*—26% of the opportunities for occurrence. This contrasts markedly with the 6.8% of such errors in the "different" group ($t = 3.62$ for percentages). Moreover, the direction and significance of the difference are not dependent on the correction factor applied to equate opportunity of occurrence. The uncorrected average number of carry-over errors was greater in the "half" group than in the "different" group (10.4 and 5.4 respectively, $t = 2.98$).

Internal evidence for habit interference in the "half" group is also strong. The group averaged over twice as many carry-over errors as errors made to unassigned letters (10.4 and 4.8 respectively, $t = 3.52$). In terms of percentage of possible errors this difference is even greater.

The most obvious interpretation of these carry-over errors would suggest that some *S*s misread or forgot the new letter assignments, and consistently responded to one or two of the old letters not now assigned. The data do not support such an interpretation. If we rank the previously assigned letters for each man of the "half" group in the order of erroneous responses to them, then a high frequency of error on the first letter or two might indicate such a tendency. Actually, the average errors fall off gradually and regularly—the means being 4.71, 3.16, 1.78, and .72 (10 possible). Seven of the 18 *S*s made one or more errors on all four of the letters. Eight more *S*s made errors on three of the four letters, and the three *S*s who confined their errors to two or fewer letters were the three with the lowest totals of such errors (0, 3, and 4).

The high frequency of carry-over errors in the "half" group might be expected to be accompanied by positive transfer in the case of previously practiced letters still assigned. Table 1 shows very little evidence for this. The tendency to omit newly assigned letters was only slightly greater than the tendency to omit letters that had been assigned on all trials. There were an equal number of each type of letter, and 13.5% error on the old letters as compared with 14.8% on the new ones. The difference is not significant ($t = .40$), and neither percent-

age of omission differs significantly from the omissions in the "different" group. Inspection of Table 1 shows that this difference was larger on Trials 7 and 8, and suggests that in this case the consideration of only Trial 6 might be misleading. Adding together the omissions on the last three trials gives a more reliable score, as well as a possible measure of more persistent transfer effects. The difference for all three trials is somewhat more in favor of positive transfer (a mean difference of 2.4 more omissions of responses to new letters, $t = 1.45$), but the difference is not significant.

Discussion

The difference between the "different" and the "half" groups with respect to negative transfer, or carry-over of the old response, poses no particular problem for theories of transfer. The momentary, external stimulus situation was the same for both groups, but the sets given by instructions to the two groups were quite different. A stimulus-response interpretation must assume that the instructions set up some sort of stimulus complex that persisted throughout learning, and that the transfer instructions set up a new complex. For the "half" group these two stimulus complexes were more similar than they were for the "different" group. Stimulus-response transfer theories have not explicitly covered this situation, but it would seem that they could easily be extended to do so.

A Gestalt theory of response production advanced by Lewin (**3, 4**) also predicts these results. Lewin denied the importance of such concepts as associative strength, and assumed instead that the intention, or determining tendency, of *S* gave rise to the response. He predicted that negative transfer occurs only when two of these tendencies are in conflict. This was the case in the "half" group. Part of the time the *S*s were correct in doing what they had previously learned to do; hence the tendency to perform in the manner previously practiced comes in conflict with the tendency to follow instructions. No such conflict occurred in the "different" condition, since the old responses were never correct.

Both types of theory agree in predicting the difference between the groups with respect to negative transfer, and they should agree again in predicting positive transfer in the "half" group on the previously practiced responses. The results, however, give no support to this prediction. Previously practiced responses were correct only slightly (and insignificantly) more often than responses to newly assigned letters. Even if further experimentation showed this difference to be significant, some explanation would still be necessary to account for its small size relative to the negative transfer.

Although a theory could undoubtedly be developed to include the results reported here, it would have to include considerably more than a consideration of associations between external stimuli and response. It would have to account for the absence of overt habit interference in the "different" group as well as for the presence of interference in the "half" group. This calls for a much more explicit theory of the effects of instructions, and probably for the assumption of some sort of mediating process. Since the tendency to respond to the old letters should have been stronger than it was to the newly assigned ones, something must have intervened. This process could have been the identification of the set of old letters. They were well

practiced and were, therefore, easily recognizable. But under the instructions given, none of these easily recognized letters called for a response. On the other hand, the letters that had never been assigned could not be as easily differentiated from the newly assigned letters, and confusion between them would be expected.

If this sort of assumption is valid, then a transfer test earlier in learning might give evidence of overt negative transfer in the "different" group. This prediction is an insecure one, since it depends on a faster initial build-up of the specific response tendency than of the recognition process. On the other hand, if the importance of the identification is increased, it should be possible to get significantly more errors of commission to letters never assigned than to previously practiced letters. This might be the case with a larger class of stimuli than letters, and an increased number of elements to which the subject must respond.

Summary

Subjects were shown a sequence of 260 letters, one letter at a time, with instructions to respond by pushing a button whenever any one of a set of eight predesignated letters appeared. After five trials with the same letters, *Ss* were divided into three groups for transfer tests. Nine *Ss* were assigned the same letters in a scrambled order ("same" group). Nine more *Ss* were assigned an entirely new set of eight letters ("different" group). Eighteen *Ss* were assigned a set of four new and four old letters ("half" group).

The results showed:

1. Complete positive transfer for the "same" group.
2. No evidence of negative transfer in the "different" group.
3. Evidence of negative transfer in the carry-over of old responses in the "half" group (previously practiced letters not assigned on the transfer trials).
4. No significant evidence for positive transfer in the "half" group on letters previously practiced and still assigned.

The results are interpreted as indicating the need for assigning a more important role in theories of transfer to sets induced by instructions.

(Received March 18, 1952)

References

1. Gagné, R. M., Baker, K. E., & Foster, H. On the relation between similarity and transfer of training in the learning of discriminative motor tasks. *Psychol. Rev.*, 1950, **57**, 67–79.
2. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
3. Lewin, K. Die psychische Tätigkeit bei der Hemmung vons Willensvorgangen und das Grundgesetz der Assoziation. *Z. Psychol.*, 1916, **77**, 212–247.
4. Lewin, K. Das Problem der Willens Messung und das Grundgesetz der Assoziation. *Psychol. Forsch.*, 1921, **1**, 191–302; 1922, **2**, 65–140.
5. Osgood, C. E. The similarity paradox in human learning: A resolution. *Psychol. Rev.*, 1949, **56**, 132–143.
6. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

NONINDEPENDENCE OF SUCCESSIVE RESPONSES IN MEASUREMENTS OF THE VISUAL THRESHOLD[1]

WILLIAM S. VERPLANCK,[2] GEORGE H. COLLIER,[3] AND JOHN W. COTTON[4]

Indiana University

Historically, psychophysics has been concerned with two kinds of functions. The first kind, that with which we shall be concerned here, includes functions that relate the probability of a response (e.g., "yes," "greater," "equal") to the spatial, temporal, and energy characteristics of stimuli; these are variously termed "psychometric," "threshold," or, as we shall call them here, "P_R" (probability of response) functions. The second type includes those that relate certain parameters (e.g., the abscissa of inflexion or 50% threshold and the *SD*) of functions of the first type to some physical dimension, again, of the stimulus. An example, in vision, of the first type is the function relating log intensity of a visual stimulus to the probability of response (**3, 18**). Examples of the second type are the visibility (**4, 31**), $\Delta I/I$ (**14, 16**), and dark-adaptation curves (**3, 13**).

Operationally, then, the great majority of investigations, both basic and applied, in the sensory and perceptual area have been concerned with first obtaining functions of the first type, that is, with determining the frequency of occurrence of some response as a function of the physical properties of the stimulating situation. The quantitative values of the various parameters of those functions have been treated as isomorphic indicators of other processes, such as mental events (sensations), events in the nervous system (nerve fibre discharges or additions of "elements"), or physical events at the receptors (absorption of certain numbers of quanta of light by the retina). The sensory, neural, photochemical, or physical processes that have been postulated to generate the observed functions have been inferred from the functions themselves, and have seldom, if ever, been available to direct or collateral observation. Limits have been placed on such inference by information obtained in other disciplines. Physiological, biochemical, and histological evidence, often scanty, has served to keep in check the variety of inferences drawn. Thus, the histological structure of the retina has placed limitations on retinal models proposed in theories of visual acuity. In vision, such men as Hecht (**15**), Wald (**30**), Crozier (**3**), and Graham (**9**), have proposed theories concerning the physical, photochemical, or neural processes responsible for observed functions.

In repeated determinations of P_R functions, it is found that the relative frequencies of a categorized response, such as "yes" or "no," to a series of presentations of stimuli varying along a single dimension under fixed experimental conditions (**7**) converge stochastically to stable values. It has, moreover, usually been assumed that, for an intertrial interval greater than the "relaxation time" of the eye, ap-

[1] The experiment reported in this paper was performed in the Department of Psychology, Indiana University, under ONR Contract N6onr-180, T. O. IV, Project 143–253, and with support from a grant by the Graduate School of Arts and Sciences. Work is continuing at Harvard University under Contract N5ori-07639.

[2] Now at Harvard University.

[3] Now at Duke University.

[4] Now at Northwestern University.

proximately .1 sec.,[5] successive trials yield a population of responses that are independent events (e.g., **6, 11, 18, 25, 26, 28**).[6] Various distribution functions (phi-gamma, Poisson, log-normal) have been proposed as relating, empirically or theoretically, the probability of response to variations in the stimulus dimension. As with the second class of function, various psychological, physiological, and physical processes have been suggested as generating the distribution functions fitted to the data, e.g., Guilford's fluctuating attention (**10**), Crozier's fluctuating thresholds of the excitable elements of the central nervous system (**3**), Hecht's quantal variability of the stimulus (**18**), and perhaps even Hull's ${}_{S}O_{R}$ (**20**).

One interesting example of a theory of the P_R function is that of Hecht (**18**). He found that under his conditions of stimulation, the amount of energy necessary to produce a "yes (I saw it)" on 60% of all the occasions on which it was presented fell in the range 2.0 to 6.0×10^{-10} ergs at the cornea, or, after correction for reflection, transmission, and absorption, 5 to 14 quanta at the retina. If it is assumed with Hecht that it is necessary for one quantum of light to be absorbed by each of 5 to 14 rods if a "yes" is to occur, then the relationship of response probability to stimulus intensity may be generated solely in terms of the quantal variability of the stimulus. On this hypothesis, Hecht derived a Poisson distribution function (predicting the absorption of n quanta by the rods) for the P_R functions which he obtained experimentally. The theoretical curves appeared to fit his data very well. One obvious corollary of this theory is that, since by the basic physical theory, the probability of n quanta arriving at the retina on any trial is independent of what may have occurred on preceding trials, then, if the time since the preceding stimulation is greater than the "relaxation time," or critical duration, the probability of a response is dependent solely on the delivery, by the stimulating flash, of the appropriate number of quanta, and is, hence, independent of preceding responses.

As we have seen, the generation of a threshold function involves the repeated presentation of stimuli, and the measurement of probabilities of response to them. Trials are defined and intertrial intervals as well. The dependent variable, relative frequency, or probability of response is one which has been employed by many behavior theorists. Operationally, a "threshold measurement" is not unlike the kind of experiment performed in studies of learning, performance, and endogenous behavior (**1, 2, 5, 8, 21, 22, 23, 27**). In view of this similarity, it seemed likely to the writers that many previous formulations must necessarily have neglected the operation of a number of variables governing behavior, and thus must be dependent for their verification on the use of a relatively limited set of values of these variables. Hence, too, such theories may be incomplete.

With this point of view, we have been concerned with the investigation of behavioral variables, such as the number of trials, intertrial intervals, and "reinforcing stimuli" on the threshold function. This paper reports one of a series of such investigations.

The first indication of the operation of behavioral variables in threshold measurements appeared in some exploratory investigations that yielded

[5] A best estimate from the literature on temporal summation.

[6] The assumption of independence is seldom, if ever, made explicitly, but it is implicit in the statistical methods employed.

data suggesting that statistical sequential *dependence* of successive responses may be found with intertrial intervals greater than the critical duration, and beyond those over which temporal summation occurs. The assumption of independence of successive responses is involved in all attempts to fit theoretically the psychophysical function by the normal probability integral, by its phi-gamma modification, or by the Poisson function, and therefore, to all theories of receptor process based on such fits. Hartline (**12**), in as yet unpublished research, found that this assumption holds for the elaboration of single nerve impulses in single fibre preparations of the *Limulus* eye, when these are stimulated some 300 successive times by a stimulus of fixed and low intensity at invariant stimulus intervals of 5 sec. or greater. At intervals of from 2 to 5 sec., the occurrence of a response may *reduce* the probability that a second flash would yield a response.

The present experiment is designed to investigate systematically the statistical independence of responses to successive visual stimulations of the human, following a procedure modeled after that of Hartline.

Apparatus and Procedure

Apparatus.—The apparatus used in this experiment was a Model III Hecht-Schlaer Adaptometer (**17**). This instrument presents binocularly a circular stimulus area subtending 3° at the retina, that is presented 7° below a red fixation point. The viewing distance is held at 25 cm. by a headrest.

A pendulum shutter provides for the presentation of the illuminated stimulus area to *S* for a period of .2 sec. The wavelength distribution of the flash is determined by a glass filter with maximum transmittance at 480 mμ. Both color temperature and intensity of the light source are held constant by the use of a fixed filament current. The luminance of the stimulus patch is continuously variable from 1.5 log μμL to 5.4 log μμL by means of a neutral optical wedge and an auxiliary neutral filter. Calibration is maintained photometrically. As the instrument was used in this experiment, flashes could be presented either at fixed mechanically determined intertrial intervals, or whenever *S* pressed a key. The apparatus was housed in a sound-deadened and light-tight room.

Subjects.—Sixteen paid Indiana University undergraduates, 12 men and 4 women, served as *Ss*. All met or exceeded Naval visual standards as tested by the Orthorater (20/20 visual acuity O.D., O.S.; no phorias; "color normal"). All were previously inexperienced in visual experimentation, so that complete control of their experimental histories was possible.[7]

Our *Ss* remained entirely uninformed of the purpose, procedures, and design of the experiment. They were instructed to maintain fixation, and to say "yes" if they saw a flash of light when the shutter operated, and "no" if they did not. When the experimental periods began, each *S* had spent 30 min. in dark-adaptation goggles and 10 min. in complete darkness. No special treatments distinguished between training, control, and experimental days, or among procedures of stimulation, except as specified later.

Procedures.—Two procedures of stimulation were employed. The first follows the *method of limits*, in which each series, ascending or descending, consisted of flashes differing successively in steps of .10 log units (LU), presented at 5-sec. intervals. Each series was terminated when three successive responses ("yes" in ascending, "no" in descending series) were made, and the next series, in the opposite direction, began with luminance .20 LU higher or lower, respectively, than the terminal luminance of the preceding series. The second procedure is the *single luminance method*. Since the problem to be investigated in this experiment is the stochastic independence of successive responses, it is necessary to eliminate the effect of changing intensities of

[7] Despite the frequent use of highly trained *Ss* in visual-threshold research, we chose to employ inexperienced persons. In view of Hecht and Shlaer's report (**17**) and of our own experience that a naive *S* will begin to make highly reliable judgments following a short period of training on threshold measurements, no loss of consistency of report need be expected. Certain advantages are attached to the use of such *Ss*. When a response is suspected to be a function of previous behavior in a situation, it is desirable to have complete control over, and records of, such behavior, which is possible only with *Ss* untrained prior to the beginning of the experiment. Further, *Ss* are less likely to divine the purpose of an experiment if their previous experience is limited, and spurious results, attributable to "suggestion," may be precluded.

stimulation on the probability of response. The method, therefore, consists of the presentation to *S* of a series of 300 flashes of constant luminance. The luminance is selected so that the probability of a "yes" in response to it by each *S* will be approximately .50. These flashes were presented in two ways. On some days, they automatically followed one another at 5-sec. intervals (A). On others, *S* was instructed to present a flash to himself whenever he was "ready" for it, by pressing a key (S).[8] In the latter procedure, the interstimulus interval was determined by *S*, who presented stimuli to himself at "optimal" times. The *S*s were allowed to take a 30-sec. rest when they wished. Insofar as possible, all sessions were run at the same time of day, on alternate days.

Experimental design.—Five preliminary sessions were given. Each such session consisted of ten ascending and ten descending series of the method of limits. From the data of each of the last four sessions, the luminance yielding a relative frequency of "yes" of .50 was determined for each *S*. The mean of these values was taken as the best estimate of *S*'s 50% threshold. These preliminary sessions provided each *S* with a standardized practice, and ensured that *S*s yielded results indicating that they followed the instructions, and that were comparable to the results usually obtained in threshold measurements. Four experimental sessions followed. Each began and ended with six ascending and six descending series on the method of limits. The middle period of these sessions was given over to the single-luminance method, in which each *S* was presented with 300 flashes; the luminance of each of these was set at the 50% threshold measured in the last four preliminary sessions. On two days, the intertrial interval was 5 sec. (A); on the other two days, the intertrial interval was controlled by *S* (S). The 16 *S*s were divided into four groups and each group was assigned one of four combinations of the two intertrial intervals over the four days of experimentation. The first group followed the order AASS, the second SSAA, the third ASAS, and the fourth SASA.

One control session was run for six randomly selected *S*s. This day differed from the experimental days only in that approximately 60 "catch" trials were introduced into the series of 300 single-luminance trials. Since it might well be argued that any results might be due to "false" responses ("yeses" not contingent on stimulation), and since the statistical treatment forbade that these be introduced on the experimental days, this final day was considered necessary. All other conditions were the same as on preceding days and *S*s were not informed of the change in procedure.

[8] This procedure aimed at duplicating the method used by Hecht in his experiments on the quantum theory of the threshold (18).

Results

Method of limits: Thresholds.—The relative frequency of "yes" responses was plotted against log luminance (log $\mu\mu$L) on normal-probability paper for each of the last four preliminary sessions on the method of limits. A straight line was visually fitted to these plots and the threshold (luminance yielding 50% "yes" responses) to the nearest tenth of a log unit was determined by interpolation.[9] For each *S* a mean threshold for the four days was determined. This luminance value is presented in the second column of Table 1. The average of the mean 50% thresholds for all *S*s was 2.68 log $\mu\mu$L, with an *SD* of .30 $\mu\mu$L.

Method of limits and single-luminance frequencies of response.—Table 1 presents the relative frequency of "yes" responses given by each *S* in each of the four experimental sessions. The mean of the distribution of differences between the .50 (predicted from the mean threshold values) and these relative frequencies was .065 and the sigma .182. A *t* test of the hypothesis that the population mean of this difference distribution is zero does not permit its rejection ($p > 5\%$ with $t = 1.57$ for 15 *df*). Within the limits of these experimental procedures, and assuming a normal distribution of thresholds, the 50% threshold, determined in the method of limits sessions, yielded, when employed in the single

[9] We do not assume that the relative frequency of "yes" versus log luminance curves employed are normal-probability ogives; this type of plot was used because it gives a convenient and reproducible graphical method for interpolating the 50% threshold values. It works.

TABLE 1

SINGLE LUMINANCE METHOD

Group	S	Preliminary Session	Day 1			Day 2			Day 3			Day 4			Control Session
		Mean 50% Threshold (log μL)	*RF* "Yes"	No. Runs	CR_R	*RF* "Yes"	No. Runs	CR_R	*RF* "Yes"	No. Runs	CR_R	*RF* "Yes"	No. Runs	CR_R	*RF* False "Yeses"
AASS	1	2.6	.75	39	5.3	.93	14	5.2	.99	2	7.0	.70	28	9.6	.03
	2	2.3	.64	22	9.7	.35	3	−0.7	.79	50	4.9	.35	65	0.8	.05
	3	2.5	.65	58	2.6	.84	32	3.9	.84	19	9.4	.85	30	4.0	
	4	2.4	.50	60	3.5	.42	63	3.5	.49	60	2.4	.42	62	2.6	.05
SSAA	1	2.7	.39	52	4.8	.30	31	8.8	.39	59	3.0	.48	54	4.8	
	2	2.7	.62	64	0.4	.61	56	3.7	.60	68	0.9	.48	70	1.1	
	3	2.9	.38	44	6.5	.43	48	6.0	.53	51	5.5	.59	45	6.6	.00
	4	2.8	.81	14	12.2	.90	18	5.5	.90	19	4.9	.88	22	5.1	
ASAS	1	2.7	.45	45	6.8	.66	33	8.8	.62	41	7.3	.58	30	10.2	
	2	2.7	.20	28	7.3	.50	42	7.6	.47	49	6.0	.81	17	11.0	.02
	3	2.2	.57	73	0.1	.36	52	4.3	.20	39	3.3	.35	58	2.6	
	4	2.7	.33	42	5.7	.19	13	12.1	.56	29	10.6	.52	19	13.0	
SASA	1	2.7	.64	67	0.5	.36	33	9.0	.37	26	8.3	.12	17	8.4	
	2	3.0	.78	36	5.4	.63	45	6.2	.99	3	4.8	.87	26	4.4	.02
	3	2.5	.34	45	5.7	.38	36	8.5	.09	15	6.0	.35	47	5.5	
	4	3.6	.57	31	10.0	.38	50	5.1	.55	41	7.7	.77	27	8.5	
Mean		2.68													.03

luminance method, on the average, 50% "yes" responses.[10]

False responses; the control session.— The last column of Table 1 presents the ratio of "yeses" to the total number of blank trials given. In two cases, this approaches 5%; in the other four, it is less. This proportion of "false" responses is small and cannot seriously affect our conclusions, although, to be sure, one would prefer that statistical necessity did not make it impossible to evaluate this basal P_R in each experimental session. This resting rate of response is somewhat analogous to an operant level.

Frequencies of response as a function of method of stimulus presentation.— The relative frequencies of "yes" responses for each *S* were averaged separately for the days when the *S*s stimulated themselves (S), and when they were stimulated automatically (A). Two corresponding distributions of differences from .50 were derived. The mean and *SD* of the A distribution were, respectively, .048 and .179, and those of the S distribution were .078 and .208. When the variances of the two distributions were compared, the null hypothesis could not be rejected ($p > 5\%$ with $F = 1.34$ for 15, 15 *df*). To test the homogeneity of means, a distribution of differences between the mean relative frequency of "yes" for the A and S

[10] This test is valid if the population sampled is normally distributed. A chi-square test of this assumption did not lead to its rejection here ($p > 5\%$ with $X^2 = 3.92$ for 2 *df*). Because of the small number of *df*, the probability of a beta-type error is large; so that we cannot accept the hypothesis of normality with any assurance. That the values sampled are *means* lends weight to the hypothesis by the Central Limit Theorem.

conditions was obtained. The mean difference was equal to .033 and the *SD* .169. A test of the hypothesis that this mean differed significantly from zero did not lead to its rejection ($p > 5\%$ with $t = 0.79$ for 15 *df*). Once again, our conclusions with respect to homogeneity of variance and homogeneity of mean relative frequencies are subject to the restriction that the tests of the normality of the distributions are not very good with so few *df* available.

Summarizing, within the limits of the present experimental procedures, and the assumption of normal distributions, the *method of limits* trials yielded 50% thresholds which on the average gave 50% "yes" responses on the *single-luminance* trials. The two intertrial procedures, self-stimulation and automatic stimulation at 5-sec. intervals, did not alter significantly the percentages of "yeses" given to the fixed luminance nor did one condition result in a more variable percentage of "yeses."

Sequential analysis: The serial correlation from trial to trial (lag 1).—In the determination of probability of response functions, a sequence of trials is given; in our case, *S*s respond with "yes (I saw it)" or "no (I did not see it)" to 300 successive presentations of a stimulus patch of invariant brightness. This procedure yields a time-ordered sequence of responses such as Yes, Yes, No, No, No, Yes, No, Yes, Yes, Yes, ... From the relative frequency of "yeses" to all 300 stimulations, an estimate of the probability of a "yes" to any given stimulation, under these conditions, can be obtained.

It has been typically assumed, implicitly or explicitly, that the successive responses, YYNNY . . ., are independent of one another, that is, that the probability of obtaining a "yes" (or a "no") on any trial does not depend on the response given to the preceding stimulus in the series. From this, it follows that the frequency of runs of responses (a run is defined as a sequence of identical responses preceded and followed by a different response) of various lengths will occur according to certain distribution laws. If significantly more runs occur than are predicted by these laws (i.e., if *S* alternates YNYNYN . . .), or fewer than predicted (YYYNNN . . .), then the assumption of independence must be rejected, and the data exhibit serial dependencies.

In order to test the hypothesis that successive responses are independent of one another, we have employed a nonparametric serial-correlation test described by Hoel (**19**).[11] The serial coefficient of correlation, R, in this test, as we have used it, is a function of the sample size, the number of "yeses" in the total number of responses, and the number of "yes" runs. R can vary between zero and plus infinity. Since the number of "yeses" varies from session to session, and since n is here relatively small for the computation of serial correlations, R's obtained from different *S*s on different days are not directly comparable. We have, therefore, used the critical ratio (CR_R) for their evaluation and comparison. This statistic has an n (0, 1) distribution. A significant positive CR_R indicates too few

[11] This test is an approximation of one developed by Wald and Wolfowitz (**29**). It is based on the cross product term

$$R = \Sigma x_i x_i - 1.$$

If +1 is assigned to each "yes" and −1 to each "no," then for 300 responses

$$R = 300 - 4(r),$$
$$E(R) = (S_1^2 - 300)/299,$$
$$\sigma R = 300 + (S_1^4 - 1196\, S_1^2 + 89400)/89102 - E^2(R),$$

where: r — number of runs, S_1 — number of "yeses" minus the number of "noes."

runs and hence a tendency to repeat a response just made; a significant negative CR_R indicates too many runs and hence a tendency to alternate. The results of this analysis are presented in Table 1, which presents both the number of runs and the CR_R for each experimental session.

Of the sixty-four CR_R's (16 *S*s over four days) calculated, only seven failed to reach the 5% level of significance. The distribution of CR_R's, averaged over all days for each *S*, gave a mean of 5.87 and a sigma of 2.25. The probability of this distribution of CR_R's[12] on the assumption of $M = 0$ and $\sigma = 1$, is very much less than 1% ($p < 1\%$ with $X^2 = 631.55$ for 15 *df*) and ($p < 1\%$ with $X^2 = ns^2/\sigma^2 = 33.75$ for 15 *df*). Serial correlation is significantly present.

A comparison between the automatic- and self-stimulation conditions was made by subtracting the average CR_R obtained with automatic stimulation at 5-sec. intervals from the average CR_R obtained with self-stimulation. The mean of this distribution was 1.22, and the sigma 2.33. A test of the hypothesis that the mean difference was zero could not be rejected ($p > 5\%$ with $t = 2.03$ for 15 *df*).

A test of the homogeneity of variance between the distribution of the mean CR_R's (over two days for each *S*) for self- and automatic-stimulation conditions again does not permit us to reject the null hypothesis ($p > 5\%$ for $F = 1.88$ with 15, 15 *df*). For further analyses the two sets of data were combined.

We may conclude that there is a significant serial correlation under both experimental conditions, and that there is no significant difference between the two conditions in the degree of serial correlation, as measured by the critical-ratio test. All but one of the CR_R's are positive, that is, runs are longer than expected on the hypothesis of a random sequence of responses. Subjects tend to repeat responses that they have just made. Serial dependencies are exhibited.

Sequential analysis: Serial correlations at lags greater than one.—The demonstration of dependent probabilities between successive responses may suggest cyclical or long-range effects of some sort. These should reveal themselves in the form of systematic changes in the index of serial correlation as the temporal interval between responses correlated increases. In order to investigate this possibility, serial-correlation coefficients at lags greater than one may be computed. That is, the correlation of the first response with the fourth, the fourth with the seventh, the seventh with the tenth, and so on (Lag 3) may be determined. These computations have been made on the present data at Lags 3, 5, 7, 9, and 20. If cyclic or regular changes in response probability occur, then the *CR* of the serial correlation coefficients should show corresponding changes. In Figure 1 are presented the mean critical ratios (over days and *S*s) plotted against lag. Up to Lag 20, the critical ratio appears to be a monotonic decreasing function of the lag. Long-range effects appear, but this analysis shows no evidence of orderly cycles.

Examination of individual records, made by plotting subsets of a single *S*'s data on a given day on binomial-

[12] Again, this test assumes a normal distribution of CR_R's. Again, there are insufficient degrees of freedom to make a reliable test of this hypothesis for the present data. However, evidence from other experiments indicates that the distribution of CR_R's is normal, with a mean different from zero and a sigma different from one in cases where significant serial correlation is present. A nonparametric sign test of the same hypothesis on the present data yields a probability between 5 and 10%.

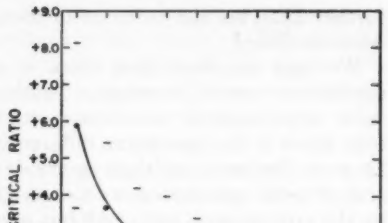

FIG. 1. Mean value of CR_B at varying lags. *SD*'s indicated by short horizontals. Visually fitted.

probability paper (**24**)[13] also reveals no effects at lags greater than 20. We do not find statistically significant lack of independence beyond about Lag 11, which, under our condition "automatic stimulation," covers about 1 min.

DISCUSSION

Successive responses were found to be dependent upon one another under all conditions of this experiment. Runs of "yeses" and "noes" of length greater than expected on the hypothesis of independence occur. If *S* reports "yes" on one flash, he is more likely to report "yes" on the next.

Furthermore, no statistically significant differences can be exhibited between either the percentage of the stimuli seen, or the dependent probabilities of successive responses when the stimulations are presented automatically, at a fixed 5-sec. interval, or when they are presented to *S* by himself when he is "ready." Further work is necessary before intertrial interval can be considered to be an irrelevant variable, since our *S*s may have stimulated themselves at a rate close to the automatic one.

Statistical nonindependence may be produced in either of two ways. First, the probability of a particular response may depend upon the identity of the previous response. Second, some as yet unidentified "state" variables may alter the probabilities of responses on trials close to one another in time—that is, there may be trends, cyclical or otherwise, in a time-ordered series, although its members are not dependent upon one another. Four categories of variables suggest themselves as possibly responsible for such effects: (*a*) stimulus variables; (*b*) nonstimulus variables independent of preceding responses, of response history, and of preceding stimulation; (*c*) nonstimulus variables independent of preceding stimulation; (*d*) nonstimulus variables dependent on preceding response history.

Category (*a*) includes variables such as fluctuations in the intensity of the stimulating light, continuous versus discontinuous presentation of the fixation point, concurrent stimulation (shutter noises, etc.), and so on. Fluctuations of the intensity of the stimulating light were controlled by our calibration and regulatory apparatus and cannot be considered a contributing variable. Extraneous stimulation was controlled by the use of a sound-deadened room, removal of relays to other rooms, and shutter-masking noises.

Category (*b*) includes such variables as physiological state (blood pressure, etc.), changes in pupillary diameter, changes in the retinal site of stimulation owing to fluctuations in fixation,

[13] This statistical tool exhibits serial dependencies in a single session's data rapidly. We used it extensively in preliminary analyses.

gross motivational fluctuations that are not a function of the immediate stimulating conditions, changes in the criterion of the *S*, and so on.

Category (*c*) includes such variables as adaptation, summation, changes in pupillary diameter produced by preceding stimulation, fluctuations in the place of stimulation dependent on preceding stimulation, etc. Adaptation is probably not a contributing variable, since its effect would be to reduce sensitivity, thereby decreasing the probability of response following a supraliminal flash, and so yielding negative serial correlations. Retinal summation over a five-sec. intertrial interval has never been reported to our knowledge.

Category (*d*) includes such variables as conditioning (i.e., an increment to the probability of response produced by the occurrence of some reinforcing event after a preceding response), peripheral or central "orientation" behavior dependent upon preceding response history, such as might result from preliminary training on the *method of limits*, motivational factors that are a function of the immediately preceding response history, and so on. At the present time, reports of experiments on the effects of many of these variables are in preparation. Other variables are still under investigation.

In any event, these results have a number of implications, not only for theories about the visual (and other) thresholds, but also for all procedures which investigate the sensitivity of the intact human. Most particularly, theories of the threshold which either assume or predict independence may need revision, and conversely, some account of the variables which produce nonindependence must be made. The results raise other questions as well: What, for example, might be done to keep *S* at peak responsiveness?

Summary and Conclusions

This experiment was performed in order to determine whether the successive responses of *S* given in measurements of his visual threshold are statistically independent of one another. This is implicit in several theories of sensory thresholds. Sixteen *S*s were presented, on four successive days, with 300 consecutive stimuli of a luminance which had been found previously to be reported 50% of the time. On two days these stimuli were presented automatically at 5-sec. intervals. On the other two, *S* presented them to himself. Critical ratios of serial-correlation coefficients were computed on the time-ordered sequences of responses at Lags 1, 3, 5, 7, 9, 11, and 20.

1. Under the conditions of this experiment the mean 50% absolute visual threshold for each subject elicited, on the average, 50% frequencies of responses on immediately subsequent days.

2. When *S* is presented with a discrete, dim flash of light, his response to it will depend not only upon the luminance of the flash, but also, among other things, on how he has responded to preceding flashes of light of the same luminance. Each response is dependent upon previous responses, or perhaps both are dependent upon a third variable which varies in time.

3. This nonindependence is exhibited with statistical significance when as many as 10 responses intervene between the correlated responses. The value of CR_R seems to be a monotonically decreasing function of the lag.

4. No statistically significant differences appeared between the data obtained under the two conditions of stimulation.

(Received for priority publication, May 19, 1952)

References

1. Bergmann, G. An empiricist schema of the psychophysical problem. *Phil. Sci.*, 1942, **9**, 72–91.
2. Bergmann, G., & Spence, K. W. The logic of psychophysical measurement. *Psychol. Rev.*, 1944, **51**, 1–25.
3. Crozier, W. The theory of the visual threshold: II. On the kinetics of adaptation. *Proc. nat. Acad. Sci., Wash.*, 1940, **26**, 334–339.
4. Crozier, W. J. On the visibility of radiation at the human fovea. *J. gen. Physiol.*, 1950, **34**, 87–136.
5. Feigl, H. Logical analysis of the psychophysical problem. *Phil. Sci.*, 1934, **1**, 420–445.
6. Fernberger, S. W. On the relation of the methods of just perceptible differences and constant stimuli. *Psychol. Monogr.*, 1912, **14**, No. 4 (Whole No. 61).
7. Graham, C. H. Psychophysics and behavior. *J. gen. Psychol.*, 1934, **10**, 299–310.
8. Graham, C. H. Behavior, perception and the psychophysical methods. *Psychol. Rev.*, 1950, **57**, 108–120.
9. Graham, C. H., & Kemp, E. H. Brightness discrimination as a function of the duration of the increment in intensity. *J. gen. Physiol.*, 1938, **21**, 635–650.
10. Guilford, J. P. "Fluctuations of attention" with weak visual stimuli. *Amer. J. Psychol.*, 1927, **38**, 534–583.
11. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
12. Hartline, H. K. Personal communication.
13. Hecht, S. The nature of foveal dark adaptation. *J. gen. Physiol.*, 1921, **4**, 113–139.
14. Hecht, S. A theory of visual intensity discrimination. *J. gen. Physiol.*, 1935, **18**, 767–789.
15. Hecht, S. Rods, cones, and the chemical basis of vision. *Physiol. Rev.* 1937, **17**, 239–290.
16. Hecht, S., Peskin, J. C., & Patt, M. Intensity discrimination in the human eye: II. The relation between $\Delta I/I$ and intensity for different parts of the spectrum. *J. gen. Physiol.*, 1938, **22**, 7–19.
17. Hecht, S., & Shlaer, S. An adaptometer for measuring human dark adaptation. *J. opt. Soc. Amer.*, 1938, **28**, 269–275.
18. Hecht, S., Shlaer, S., & Pirenne, M. H. Energy, quanta, and vision. *J. gen. Physiol.*, 1942, **25**, 819–840.
19. Hoel, P. G. *Introduction to mathematical statistics.* New York: Wiley, 1947.
20. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
21. Johnson, H. M. Did Fechner measure 'introspective' sensations? *Psychol. Rev.*, 1929, **36**, 257–284.
22. Johnson, H. M. Are psychophysical problems genuine or spurious? *Amer. J. Psychol.*, 1945, **58**, 189–211.
23. Kantor, J. R. Can the psychophysical experiment reconcile introspectionists and objectivists? *Amer. J. Psychol.*, 1922, **33**, 481–510.
24. Mosteller, F., & Tukey, J. W. The uses and usefulness of binomial probability paper. *J. Amer. statis. Ass.*, 1949, **44**, 174–212.
25. Mueller, C. G. Quantum concepts on visual intensity-discrimination. *Amer. J. Psychol.*, 1950, **63**, 92–100.
26. Stevens, S. S., Morgan, C. T., & Volkmann, J. Theory of the neural quantum in the discrimination of loudness and pitch. *Amer. J. Psychol.*, 1941, **54**, 315–335.
27. Tinbergen, N. *The study of instinct.* Oxford: Oxford Univer. Press, 1951.
28. Urban, F. M. The application of statistical methods to the problems of psychophysics. In Lightmer Witmer (Ed.), *Experimental studies in psychology and pedagogy.* Philadelphia: The Psychological Clinic Press, 1908.
29. Wald, A., & Wolfowitz, J. An exact test for randomness in the nonparametric case based on serial correlation. *Ann. math. Statis.*, 1945, **14**, 378–388.
30. Wald, G. Area and visual threshold. *J. gen. Physiol.*, 1938, **21**, 269–287.
31. Wald, G. Human vision and the spectrum. *Science*, 1945, **101**, 653–658.

FACTORS RELATED TO SPEED AND ACCURACY OF TACTUAL DISCRIMINATION[1]

T. R. AUSTIN AND R. B. SLEIGHT

The Johns Hopkins University

In a previous experiment (1), speed and accuracy of tactual discrimination were tested for 43 figures. These included the upper-case letters A through Z, the numerals 2 through 9, and nine geometric forms. Results showed that 25 figures—or approximately 60% of the total number of stimuli—met or exceeded a 90% accuracy criterion. These results pertained to a solid set of figures when finger-tip movement was permitted. One implication of the study seemed to be that such tactual cues might be applied to instrument panels, dashboards, machinery switches, etc.

The experiment described above was exploratory in nature. Its main purpose was to determine, in fact, if the finger tips *could* be employed in discriminating these types of figures. The positive evidence warranted further investigation, and this paper reports the results of a subsequent study.

If tactual cues are to be utilized, it is appropriate to determine whether additional factors play a part in *S*'s ability to discriminate. The purpose of this experiment was, therefore, to determine if speed and accuracy of tactual discrimination were related to or affected by differences in sex, handedness, fingers employed in making the discriminations, learning, pressure exerted by the finger tip, and levels of subjective confidence.

[1] This research was supported in part by Contract N5-ori-166, Task Order I, between the Systems Coordination Division, Naval Research Laboratory, Office of Naval Research, and The Johns Hopkins University. This is Report No. 166-I-155, Project Designation No. NR-784-001, under that contract.

Apparatus

The set of figures tested was selected from the solid group of figures used previously. These figures were cut from ¼-in. Masonite sheeting and were the maximum size which could be inscribed in ½-in. circles. Stroke width was held constant at 2 mm. Some slight alterations from common forms were made in the figures when such alterations enabled the over-all area of the figures to be increased within the ½-in. circles. The figures tested were the letters C, D, E, G, I, J, K, L, O, P, Q, T, U, V, W, X, and Y, the numerals 2, 7, and 9, and the following geometric forms: diamond, right triangle, rectangle, crescent, and half-circle. These figures are the ones which met or exceeded a 90% accuracy criterion in the first study.

The figures were mounted on a plywood disk which could be revolved 360° in the horizontal plane to expose one figure at a time through a slot in the handrest. The pivot pin in the center of the disk was secured to a lever which was displaced in the vertical plane by *S*'s pressure on the figure. The movement of the lever was geared to a synchro-generator, and the latter was electrically wired to a vacuum tube voltmeter. By means of this arrangement, the deflection of the voltmeter was linearly proportional to the amount of pressure exerted on the figure by the finger tip, and the voltage could easily be transformed into pressure units.

The slightest pressure by *S*'s finger closed a microswitch which activated a clock in a voice reaction timer. When *S* responded, the timer was stopped by means of a throat microphone. The apparatus produced highly accurate measures of reaction time and pressure exerted by the finger tip at the instant of discrimination.[2]

Procedure

The *S*s were seated at a table on which the equipment was placed. The distance from *S* to the table was adjusted so that all discriminations were made with partially extended arms. The *S*s were told that the task was to discriminate letters, numerals, and geometric forms, but were

[2] The authors wish to express their thanks and appreciation to Mr. T. W. Hollis, Jr., who designed and built the tactual discrimination apparatus.

TABLE 1

ORDER OF FINGERS TESTED FOR 16 *Ss* IN A MULTIPLE LATIN-SQUARE DESIGN

After the four initial trials, each *S* was tested in four additional trials with the finger employed first in his sequence. Roman numerals in the matrix of this table correspond to the fingers employed in each specific trial.

S		Trial				
Sex	Handedness	1	2	3	4	5–8
M	Right	III IV I II	I II IV III	IV III II I	II I III IV	III IV I II
M	Left	III IV I II	II I III IV	IV III II I	I II IV III	III IV I II
F	Right	II III I IV	IV I II III	III II IV I	I IV III II	II III I IV
F	Left	II I III IV	I III IV II	III IV II I	IV II I III	II I III IV

not told the number or relative frequency of each type. Instructions were as follows:

"Place your hand in position and at the signal from the experimenter lower the specified finger on the figure. You may move your finger at will. Respond as soon as you think you are sure you know what the figure is. Speed is important but accuracy is of primary importance, so make sure of your discrimination before responding. After each response, indicate the confidence of your discrimination on a five-point scale." (The five-point scale was thoroughly explained by *E*.) "Remember, accuracy should be your primary aim in each discrimination."

Accuracy was emphasized in the instructions as an attempt to guard against *Ss*' responding too rapidly and not as accurately as they might. Reaction time, exerted pressure, and level of confidence were recorded for each response. The *Ss* were not blindfolded but were instructed not to look at the apparatus at any time throughout the experiment. They followed this instruction without exception. None of the *Ss* had previous experience with this type of tactual discrimination.

Sixteen *Ss* (eight men, four right-handed and four left-handed; and eight women, four right-handed and four left-handed) were tested on their four *right-hand* fingers (thumbs excluded). Handedness was established by asking each *S* questions from the Purdue Speech Clinic Laterality Index. The order of testing fingers of each *S* was according to a multiple latin-square design as illustrated in Table 1. It is seen that none of the finger sequences was repeated.

After each *S* had four trials, one with each finger in the specified latin-square sequence, he was given four additional trials with the finger which was tested *first* in each sequence. These additional trials were incorporated to enable an analysis of learning trends to be made. Each trial consisted of 25 discriminations, one for each figure in a random order. A different order was used for each *S* for each trial.

RESULTS

Accuracy.—Accuracy and reaction time were held to be the most important factors in the investigation. The increase in accuracy over the eight-trial period is shown in Fig. 1. Mean per cent accuracy increased from 87.36% (Trial 1) to 100% (Trials 7 and 8). Each mean is the average of 400 discriminations (16 *Ss*, 25 discrimnations per *S*).

Reaction time.—In Fig. 2, mean reaction time is plotted as a function of successive trials. It is seen that during the eight-trial period, mean reac-

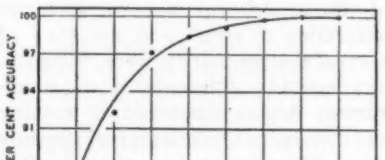

FIG. 1. Increase of mean per cent accuracy during the eight experimental trials. Each point is the mean of 400 discriminations (16 *Ss*, 25 discriminations per *S*).

tion time decreased from 9.11 sec. (Trial 1) to 3.52 sec. (Trial 8). Corresponding to the steady decrease in reaction time was the variability of reaction-time scores. The *SD*'s decreased from 2.29 on Trial 1 to 1.59 on Trial 8.

Confidence.—Although confidence never reached a mean five level (perfect confidence), it increased throughout the range of trials. Figure 3 expresses this relationship, and it may be seen that mean confidence increased from 3.99 on Trial 1 to 4.92 on Trials 7 and 8.

Pressure.—The pressure exerted by the finger tip at the time of each discrimination was read directly from the voltmeter by means of a special scale calibrated in ounces. Pressure was recorded by scoring a "1" for all discriminations effected with zero to 3 oz. of pressure, "2" for the 3- to 6-oz. interval, etc. In only 30 responses from a total of 3200—less than 1%—was the zero- to 3-oz. interval exceeded.

Sex, handedness, and fingers.—These factors obviously did not contribute a major source of *error* variance since accuracy reached the 100% level for all Ss tested. In addition, a simple analysis of variance was performed on the *reaction-time* scores of the first four

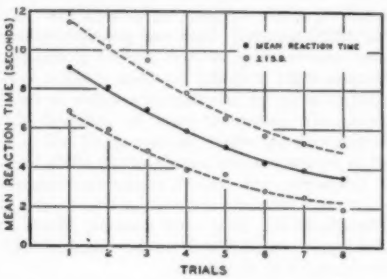

FIG. 2. Decrease of mean reaction time as a function of successive trials. Each point is the mean of 400 discriminations.

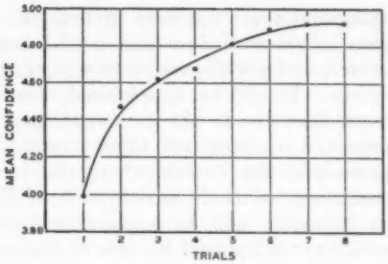

FIG. 3. Increase of mean confidence over the eight-trial period. Each point is the mean of 400 discriminations.

trials of each *S* to determine whether sex, handedness, or fingers differed significantly in terms of reaction time. Resulting variances for all three factors were small enough to be neglected. Both accuracy and reaction time data, therefore, strongly suggest that there are no differences between males and females, left- and right-handed people, or any of the four right-hand fingers in ability to tactually discriminate these figures.

Subject analysis.—All the above results have been presented in terms of trial-by-trial analysis in order to express the relationships between speed and accuracy and the various other factors over a series of repeated trials. It remains to compare speed, accuracy, and confidence of the 16 individual Ss.

Three product-moment correlations were computed: between reaction time and per cent accuracy, between reaction time and confidence, and between per cent accuracy and confidence. For each correlation there were 16 paired scores, one pair for each *S*. Each score was the mean of the 25 separate discriminations for each of the first four trials taken together. Therefore, a score used in calculating a correlation was the average of 100 discriminations. Only the first four trials were employed because beyond

this point there was very little dispersion in reaction-time and confidence scores, and no dispersion in accuracy scores. It will be appreciated, however, that due to the combination of repeated trials on the same stimulus materials, the correlated scores are comprised of both initial level of *S* performance and subsequent values which are influenced by rate of learning. These correlations were computed to answer the following questions: (*a*) Did *S*s who responded accurately also respond quickly; (*b*) did *S*s who responded quickly do so with high confidence; and (*c*) did *S*s who responded accurately do so with high confidence?

The product-moment r between mean reaction time and mean per cent accuracy was $-.70$. This r shows a significant relationship between short reaction time and high accuracy, and reciprocally, long reaction time and low accuracy, at beyond the 1% level of confidence.

For the correlations involving confidence scores, only the mean number of "five" confidence reports was employed. In other words, mean perfect confidence was correlated with reaction time and per cent accuracy.

When the correlation between mean reaction time and mean confidence was calculated, the data yielded an r of $-.94$, indicating to a marked degree that high confidence was accompanied by short reaction-time scores, and low confidence was accompanied by long reaction-time scores.

The correlation between mean confidence and mean per cent accuracy resulted in an r of .47 (significant only at the 10% level). Prediction of accuracy from confidence scores, therefore, would be less dependable than prediction of reaction time from confidence scores.

Discussion

Learning.—The learning that resulted from the repeated trials is apparent from an analysis of speed and accuracy scores. Both of these measures indicated that a great deal of learning did take place during the experimental sessions. Mean reaction time for the 25 figures on the first trial was shown to be 9.11 sec., and the corresponding mean accuracy was 87.36%. Seven trials later, mean reaction time had decreased to 3.52 sec., and mean accuracy had reached the 100% level. Accompanying these trends was a steady decrease in variability of the reaction-time scores.

Equipment operation.—When the speed and accuracy data are considered along with the evidence that sex, fingers, and handedness are nonsignificant factors in these discriminations, several implications of importance for selection and training problems may be seen. Any piece of equipment employing tactual cues would likely have to be standardized, i.e., be located in the same position, made of the same design and materials, etc., as demanded by production methods. These results indicate that males and females, and left- and right-handed operators are equally proficient in the utilization of the cues. Therefore, for example, the tactual array might always be situated to the right-hand side of *any* operator. Further, the results indicate that all four right-hand fingers may be employed with equal efficiency. This would suggest that all four fingers could be used on a battery of push buttons, thus enhancing the versatility of the array. The most important aspect of the results, however, is the fact that all persons can reach the 100% accuracy level and greatly diminish their reaction time in short periods of training. Because eight trials did not seem to allow sufficient practice for reaction-time scores to reach asymptotic values, it may be assumed that further training would increase speed still more while accuracy would remain at the 100% level.

Confidence.—In general, confidence increased as accuracy increased and reaction time decreased throughout the eight trial sessions. Such an increase in confidence is to be expected with familiarity with the figures.

An unexpected result was the high correlation between confidence and reaction time and the considerably lower correlation between confidence and accuracy in the subject analysis.

What *S* is willing to state in the way of confidence is more highly correlated with how fast he is willing to state it than the accuracy with which the discrimination is made. Regardless of the low correlation between accuracy and confidence, however, each of the 16 *S*s reached the 100% level of accuracy by his seventh trial. It therefore seems that confidence was not a basic criterion for these discriminations, especially when accuracy was the most strongly emphasized factor.

Pressure.—This measure was the most highly consistent and possibly one of the most important of those recorded. Approximately 99% of the discriminations were made by exerting between zero and 3 oz. of pressure. This consistency is important when considered in terms of equipment design. If tactual cues should be used on push buttons, and if the push buttons are to be spring loaded, then the finger tip must depress the spring as well as afford a discrimination. The consistency of pressure measurements in this experiment, both for different *S*s and different fingers, suggests that a common standardized spring might possibly be built into all push buttons.

Summary and Conclusions

The purpose of this experiment was to determine if speed and accuracy of tactual discrimination were related to or affected by differences in sex, handedness, fingers employed in making the discriminations, learning, pressure exerted by the finger tips, and levels of subjective confidence.

The stimulus figures utilized had all met or exceeded a 90% accuracy criterion in a previous experiment. The figures used were 17 upper-case letters, three numerals, and five geometric forms.

Sixteen *S*s (eight men, four right-handed and four left-handed; and eight women, four right-handed and four left-handed) were tested.

The results of the experiment seem to warrant the following conclusions:

1. A great deal of learning took place during a short training session as shown by a marked increase in accuracy and decrease in reaction time.

2. Levels of subjective confidence increased as accuracy increased and reaction time decreased throughout the eight-trial period.

3. The product-moment correlation between mean per cent accuracy and mean reaction time was −.70; between mean confidence and mean reaction time −.94; and between mean confidence and mean per cent accuracy .47.

4. In approximately 99% of the discriminations, pressures from zero to 3 oz. were exerted in making the discriminations.

5. No significant differences resulted from discriminations made by left- versus right-handed *S*s, males versus females, or any of the four right-hand fingers.

(Received for priority publication June 20, 1952)

References

1. Austin, T. R., & Sleight, R. B. Accuracy of tactual discrimination of letters, numerals, and geometric forms. *J. exp. Psychol.*, 1952, **43**, 239–247.

FUNCTIONAL FIXEDNESS AS RELATED TO PROBLEM SOLVING: A REPETITION OF THREE EXPERIMENTS [1]

ROBERT E. ADAMSON

Stanford University

The study of problem solving and thinking has been retarded by the lack of agreed-upon theoretical concepts supported by adequate data from experiments. As a part of a larger program concerned with these matters, some of the more promising hypotheses have been assembled, and preliminary experiments undertaken to repeat the demonstrations upon which these hypotheses rest.

One inviting hypothesis is that problem solving may in some instances be delayed through the "functional fixedness" of solution objects. That is, owing to his previous use of the object in a function dissimilar to that demanded by the present problem, *S* is inhibited in discovering the appropriate new use of the object. This hypothesis was proposed by Duncker (3), who designed ingenious experiments to support it, but carried the experiments through with but 14 *Ss* and under poorly specified experimental conditions. It seemed wise, therefore, to repeat some of his experiments both to substantiate his results, if possible, and to ascertain the efficacy of the problems for use in further investigations. The success of Birch and Rabinowitz (1) in demonstrating functional fixedness in a related experiment encouraged us to hope for positive results.

[1] This experiment was the first in a series of studies of problem solving being done under Project NR 150–104 and supported by Contract Nonr 225 (02) between Stanford University and the Office of Naval Research. The work was done and the present report prepared under the supervision of Dr. Donald W. Taylor. Work on the contract is under the general direction of Dr. E. R. Hilgard.

Procedure

Subjects.—All *Ss* taking part in this study were college students from elementary psychology classes. There were 57 *Ss*, of whom 35 were men, 22 women. Twenty-nine *Ss* were assigned to the experimental group, 28 to the control group. All *Ss* were of proximate ages and had been exposed to little experimentation.

Problems.—Duncker's "box," "gimlet," and "paperclip" problems were presented to each *S* in the order named. In the first of these, the box problem, *S*'s task is to mount three candles vertically on a screen, at a height of about 5 ft., using to accomplish this task any of a large number of objects which are lying before *S* on a table. Among these objects are three pasteboard boxes of varying sizes, five matches, and five thumbtacks, the crucial objects for solution of the problem. The solution is to mount one candle on each box by melting wax on the box and sticking the candle to it, then to tack the boxes to the screen.

The gimlet problem involves suspending three cords from a board attached to an overhead beam. Among the variety of objects available are two screw-hooks and the gimlet itself, the objects from which the cords may be hung.

The paperclip problem consists of first attaching four small black cardboard squares to a large white square, then hanging the large square from an eyelet screwed into the aforementioned beam. Included among the objects lying before *S* on the table are a number of paperclips. These may be used to attach the small squares to the large one, and one of them, when bent to form a hook, will serve to hang the large square from the eyelet.

Design.—The experimental and control groups were given the same problems to solve. For the experimental group, however, at least one of the solution objects was "burdened" with a prior function in each problem. Thus, the candles, matches, and tacks for the box problem were placed in the three boxes before they were given to *S*. Hence, the boxes had for their initial function that of containing, whereas in their solution function they had to be used as supports or

platforms. Similarly the gimlet initially had to be used to start holes for the screw-hooks, and in the paperclip problem, the four black squares had to be attached to the white one with paperclips. Duncker referred to the experimental group as the "after pre-utilization" group.

The control group was given the problems without any pre-utilization. In the case of the box problem, the empty boxes were placed on the table at varying distances from the other crucial solution objects. Holes into which the screw-hooks and the gimlet could be screwed were already drilled into the beam in the case of the gimlet problem; the four black squares were stapled to the white one in the paperclip problem. Thus, none of the crucial objects was used with a function prior to its use as a solution object.

Solution scores were taken as one possible measure of functional fixedness, and time-to-solution constituted another measure. A maximum time of 20 min. was allowed for solution of each of the problems.

Results

Box problem.—The results of the box problem, presented in Table 1, confirm Duncker's finding that functional fixedness results from pre-utilization. The performance of the experimental group was markedly inferior to that of the control with respect both to the number of solutions obtained and the time required to reach solution. Prior usage of the boxes as containers inhibited their being used as platforms.

The chi-square value comparing the two groups on the time score was obtained by using as a cutting point the median time-to-solution of the combined groups. All cases for which there was no solution were assigned to the above-median category. With 1 *df*, each of the chi squares was highly significant. Since the direction of the difference was predicted, a one-tail test of significance was employed for both this and the following two problems.

TABLE 1

Box Problem

Group	n*	Number Solving	Time-to-Solution*
Exper.	29	12 (41%)	7 (24%)
Control	28	24 (86%)	22 (78%)
		$\chi^2 = 12.0$ $p = .001$	$\chi^2 = 14.8$ $p = .001$

* Number below median of combined group.

TABLE 2

Gimlet and Paperclip Problems

Problem and Group	n	Mean Time-to-Solution (Sec.)	*SD*	t'	p*
Gimlet					
Exper.	26	246.6	124.7	3.71	.001
Control	28	144.0	67.7		
Paperclip					
Exper.	29	107.9	96.0	2.38	.01
Control	28	63.0	31.5		

* Single-tail test.

Gimlet problem.—Since only three *S*s failed to solve this problem, all from the experimental group, the solution score could not demonstrate a difference between the experimental and control groups. Accordingly, only the results from the time measure are given in Table 2. The three *S*s failing to reach solution were not considered in the analysis of the data, thus reducing the total n to 54 for this experiment.

Since, as shown in Table 2, the variances for the two groups are not homogeneous, the use of t as a test of significance was inappropriate. Instead, t' was employed.[2] The highly significant difference obtained shows clearly the presence of functional fixedness.

Paperclip problem.—The results from the paperclip problem are also shown in Table 2. Since all *S*s were able to solve this problem, only time scores are given. As in the first two problems, pre-utilization of the solu-

[2] This technique was suggested by Dr. Quinn McNemar. Instead of utilizing one estimate of the population variance, t' incorporates the estimated variances from two populations (2). It is, in consequence, useful in such a situation as the present one, replacing t which assumes homogeneity of variance.

tion objects with a function different from that demanded by the problem resulted in significantly poorer performance by the experimental group.

Reliability of individual performance.—An analysis was made of the performance of the experimental group to determine whether individual achievement on one of the three problems was significantly related to achievement on either of the other two. Chi square was used to test whether individuals scoring below the median in time-to-solution for one of two problems also showed a significant tendency to score below the median for the other. (Since only 12 of 29 *Ss* solved the box problem, a median time-to-solution could not be obtained; instead, the distribution was dichotomized in terms of solution or no-solution.) A relation significant at the .05 level was found between the box problem and the paperclip problem. Neither of the other chi squares was significant. Clearly, achievement on a single problem involving pre-utilization is not a reliable measure of individual susceptibility to functional fixedness.

Discussion

Duncker's study (3), involving these three experiments, used two measures of performance: number of presolutions, and number of solutions. In the present study, the number of presolutions was discarded as a measure, because it was found to be overly dependent upon the subjective judgment of *E*. Number of solutions proved to be a satisfactory measure for only the box problem. Since all *Ss* solved the paperclip problem, and all but three solved the gimlet problem, no difference between the experimental and the control groups could be revealed by this measure. It would appear that the *Ss* in the present study were more able than those employed by Duncker.

Although the measures previously used by Duncker failed to show functional fixedness in two of the three present experiments, a new measure, time-to-solution, gave positive results in all three experiments. Essentially, then, the present results confirm those obtained by Duncker.

The results of Duncker, of Birch and Rabinowitz, and those obtained in this study afford convincing proof of the existence of functional fixedness. The reality of this phenomenon having been established, two lines of investigation are of immediate interest: (*a*) determination of those conditions which influence the occurrence of functional fixedness, and (*b*) exploration of its relation to other kinds of set in problem solving. A study now nearing completion involves both of these lines of experimentation.

Summary

1. Three of Duncker's experiments on functional fixedness were repeated in this study. Fifty-seven *Ss* were used, 29 serving as the experimental and 28 as the control group. Both groups were given the "box," "gimlet," and "paperclip" problems in that order. Experimental *Ss* were given each problem after first having used the solution objects for that problem in a function dissimilar to that demanded for solution. Control *Ss* were given the problems without such pre-utilization.

2. Two measures of performance were used: number of solutions, and time-to-solution. The former measure discriminated between the experimental and control groups on only the box problem; the latter measure gave highly significant differences in the expected direction for all three problems.

3. Functional fixedness was shown to result from the pre-utilization of solution objects. Duncker's results were confirmed in a study using a larger n and having more carefully specified experimental conditions.

(Received for priority publication June 26, 1952)

References

1. Birch, H. G., & Rabinowitz, H. S. The negative effect of previous experience on productive thinking. *J. exp. Psychol.*, 1951, **41**, 121–125.
2. Cochran, W. G., & Cox, G. M. *Experimental design.* New York: John Wiley & Sons, 1950.
3. Duncker, K. On problem-solving. *Psychol. Monogr.*, 1945, **58**, No. 5 (Whole No. 270).

SIMULTANEOUS VISION AND AUDITION: THE DETECTION OF ELEMENTS MISSING FROM OVERLEARNED SEQUENCES [1]

G. H. MOWBRAY

Cambridge University

As transportation by land, sea, and air has become swifter and more complex, the necessity for communicating more information in shorter periods of time has arisen. This has led quite naturally to a critical evaluation of existing channels of communication and an extended search for new mediums to exploit. As an outgrowth of the mounting interest displayed in the development of visual communication systems, Cheatham (**2**) has surveyed the pertinent psychological literature on vision and audition in order to compare these two sense modalities as communication channels. Harris (**4**), in an earlier report, has considered at some length the relations between vision and audition from the standpoint of a dimensional comparison rather than a direct one based on a specific laboratory experiment. The numerous attempts to discover whether auditory or visual presentation of information is more efficient have recently been reviewed by Day and Beach (**3**). These authors abstracted 11 generalizations from the literature which specify the conditions under which visual or auditory presentation is superior. Except for a discussion of intersensory facilitation by Harris and a consideration of the combined visual and auditory presentation of the same material by Day and Beach, all of these papers are exclusively concerned with the two sense modalities operating in isolation. While these are necessary and important steps to be taken, the picture can by no means be complete without data bearing on the question of the relative efficiency of the two sensory systems when they are simultaneously stimulated by conflicting material.

The nearest approach to this problem is provided by the classical experiments on division of attention which are reviewed by Woodworth (**8**). In almost every case, the simultaneous tasks provided have one sensory and one motor component although one or two are reported that combine a tactile and a visual discrimination or two simultaneous visual discriminations. The orientation of all of this research has been toward discovering if successful division of attention can be said to occur. No systematic attempt has been made to compare the relative performances of any two sensory modalities operating simultaneously.

Previous research has shown that, for single sensory channel operation, the type of material utilized for comparative purposes exerts a major influence in determining which modality will dominate. Pohlmann (**7**) found that the visual sense was superior

[1] This research was supported in part by Contract N5-ori-166, Task Order I between the Systems Coordination Division, Naval Research Laboratory, Office of Naval Research, and The Johns Hopkins University. This is Report No. 166-I-121, Project Designation No. NR-507-470, under that contract. The actual experimental work was done at The Psychological Laboratory, University of Cambridge, Cambridge, England and the author is grateful to Professor Sir Frederic Bartlett for the facilities and assistance provided at that laboratory. Thanks are also due to Dr. W. R. Garner of The Johns Hopkins University for his helpful criticisms and suggestions.

when nonsense syllables were presented but that the auditory mode dominated when meaningful material was used for testing. Carver (1) reported vision better than audition for difficult words but the exact opposite for easy words, and Larsen and Feder (5) partially confirmed this result when they found difficult but meaningful factual material was handled better visually. However, with easier factual material there was no difference between the sense modalities in question. In general, then, if the material used in testing is difficult or unfamiliar, the most effective manner of presenting it is probably by way of the visual channel. On the other hand, easy material may best be handled through audition, although there may well be no difference between the senses in this latter case.

Purpose of Experiment

Since difficulty of material is apparently an important parameter in comparative studies of sensory channels operating independently of each other, it may also be an important consideration when the modalities are operating simultaneously. The purpose of this experiment, therefore, was to test the relative efficiency of vision and audition under conditions of simultaneous stimulation when "simultaneous" is understood to refer to the occurrence of a continuous series of events within the same span of time.

By providing perceptual tasks of a continuous nature (as opposed to isolated, momentary perceptions) two things were accomplished. First, the necessity for and the technical difficulty involved in temporally equating the visual and auditory stimuli were in some measure obviated; and, second, a much closer approximation to a natural *modus operandi* was obtained. The task that was selected as most adequately conforming to experimental demands was one which required *Ss* to detect and report elements randomly deleted from familiar sequences. Series of symbols that were grossly overlearned were considered to provide the necessary, familiar substratum as well as to permit the selection of varying levels of difficulty on the basis of the amount of overlearning within a particular sequence. Since two sequences are already established in most literate, English-speaking adults which can reasonably be said to be overlearned by differing amounts, these were utilized for this experiment. The sequences in question are English alphabets and Arabic numerals.

Method

Materials.—The visual sequences consisted of thin white strips of cardboard on which were printed English alphabets or Arabic numeral sequences from 1 to 20. All sequences had either 2, 3, 4, 5, or 6 elements missing as determined by a random-selection process. Samples of a typical alphabet and a typical numeral sequence are presented in Fig. 1. Both of these samples have three elements deleted. In a given sequence (either numeral or letter) any symbol could be omitted as determined by chance. The auditory stimuli were similar sequences recorded on magnetic tape.

Procedure.—The *Ss* were seated in a darkened room facing a white square of cardboard to which was affixed the visual stimulus. They were told that following a ready signal they would be shown a numeral sequence (or an alphabet, depending on the condition being tested) for a brief period of time, and that they were to read it through once from left to right as rapidly as they could and then write down those numerals (or letters) which they had detected as missing from the sequence. They were never told what the

1 2 3 5 6 7 8 10 11 13 14 15 16 17 18 19 20

A B C E F G H I J K L N P Q R S T U V W X Y Z

Fig. 1. A sample of an alphabet and a numeral sequence. Each sequence has three elements deleted.

range of possibilities was—whether there would be none missing or ten missing or all missing. This was the procedure for all nonsimultaneous visual trials. Exposure of the stimulus was accomplished by flashing a beam of light on the cardboard stimulus strip by means of a hand-operated shutter mechanism on a slide projector. The exposure time for all visual trials was 6 sec., which was barely long enough for the fastest *S*s to skim through the sequences once. The auditory trials, similarly, were of 6-sec. duration, and all sequences were preceded by a warning note recorded on the magnetic tape 3 sec. before the beginning of the series.

Practice trials.—The *S*s were given practice on both the auditory and the visual trials with both numerals and alphabets until no trial-to-trial improvements were noted in the number of detection errors committed. Two or three experimental sessions were all that were required for *S*s to reach a performance plateau, and control checks made throughout the course of the experiments showed no further improvements. All practice trials consisted of the visual and auditory conditions being presented separately.

Nonsimultaneous trials.—Following the practice trials, *S*s were tested on all four nonsimultaneous conditions (i.e., visual numeral, auditory numeral, visual alphabet, and auditory alphabet) to obtain response scores in terms of the average number of detection errors per sequence as a function of the number of elements missing in order to define the basic level of single sensory channel operation for each *S*. Two, and in some cases three, experimental sessions were devoted to collecting scores for each of the four stimulus conditions just mentioned. One experimental session consisted of 50 sequences—10 each with 2, 3, 4, 5, and 6 elements missing. There was a random order of presentation of sequences within an experimental session.

Simultaneous trials.—The simultaneous trials included all combinations of visual and auditory sequences that were possible with the restriction of double-channel stimulation imposed—visual numeral and auditory numeral, visual numeral and auditory alphabet, visual alphabet and auditory alphabet, and visual alphabet and auditory numeral. Each experimental session consisted of 50 sequences presented to one modality and 50 sequences presented to the other. Within each experimental session all possible combinations of numbers of elements missing in one sequence were presented with all others an equal number of times. That is to say, the auditory sequences were composed of ten with two elements missing, ten with three elements missing, ten with four elements missing, etc.; the visual sequences were similarly constructed. The auditory sequences with two elements missing and the visual sequences with two elements missing were presented together the same number of times as all other combinations. Since there were 25 combinations, each was presented twice in any one experimental session. When the simultaneous presentations consisted of visual numerals and auditory numerals or visual alphabets and auditory alphabets, the problem of the coincidence of deleted elements presented itself. Since for any given sequence the missing elements were selected randomly, it was possible, on a chance basis, for the same symbol to be missing in both the auditory and visual sequence. For half of the simultaneous trials *S*s recorded their visual results first and auditory results second. On the other half of the trials this procedure was reversed. Two experimental sessions were devoted to each of the four simultaneous conditions stated previously.

Subjects.—Ten enlisted men in the British Navy, whose ages were between 17 and 29, served as *S*s. Most *S*s contributed 600 nonsimultaneous observations, but some contributed only 400 since it was found that the experimental time could be reduced without seriously affecting the precision of the results. All *S*s contributed 400 simultaneous observations on 800 sequences.

Results

Types of error.—The nature of the perceptual task made it possible to commit two types of detection errors. The first of these consists of the failure to report elements which were missing from the sequence; these have been termed *errors of omission.* The second consists of reports of elements missing which were, in fact, present in the series; these have been termed *errors of commission.*

Nonsimultaneous trials.—Since the data from the simultaneous trials have been analyzed with respect to *S*'s performance when only one sensory channel was being stimulated at a time, the first results of importance are those relating to this latter condition. Figure 2 shows the average number of errors per sequence as a function of the number of elements missing for all nonsimultaneous presentations and for all *S*s. For greater clarity the

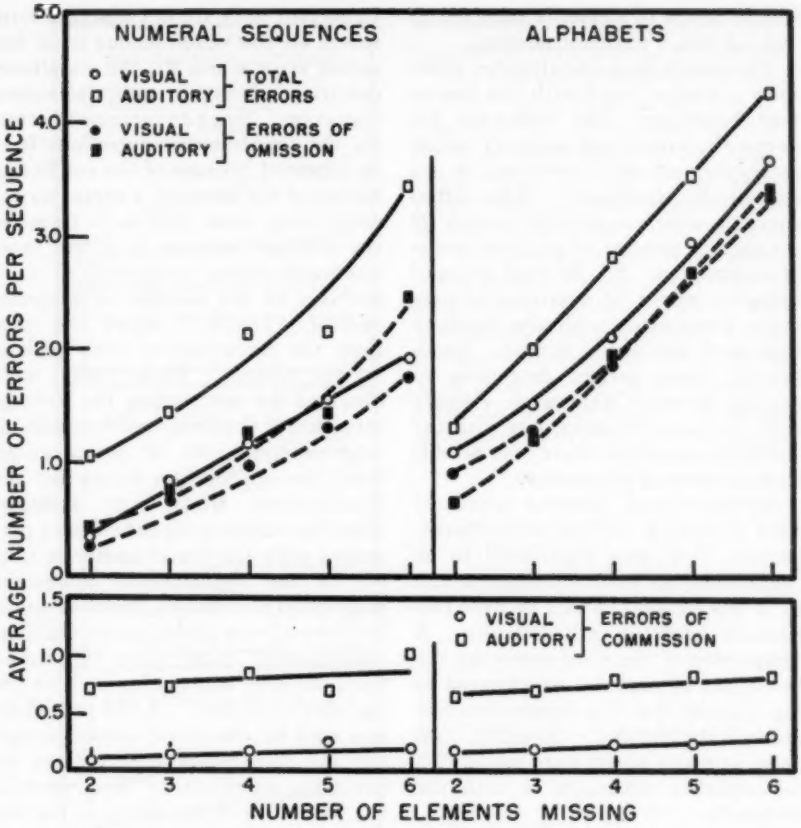

FIG. 2. The average number of errors per sequence as a function of the number of elements missing for all nonsimultaneous conditions

numeral series and alphabet series are presented on different coordinates. The three groups of functions represent total errors, errors of omission, and errors of commission for both visual and auditory sequences. In considering the numeral series first, what is immediately evident is the considerably greater number of total errors committed with the auditory stimuli. When the total errors are broken down into errors of commission and errors of omission, it is seen that this discrepancy is largely caused by the fact that about five times more errors of commission were made aurally than were made visually. However, the number of errors of omission is significantly greater for the auditory sequences than for the visual sequences when the results are averaged over the five conditions of numbers of elements missing and subjected to a t test for correlated scores ($p < .001$). Thus, more of both types of errors were made aurally than visually, but the greatest difference between the modalities for the nonsimultaneous numeral series

was in terms of elements reported as missing which were not missing.

The results from the alphabet series show a similar trend with one important exception. The difference between the visual and auditory series, as regards errors of omission, is not statistically significant. What differences do exist are entirely caused by the greater number of auditory errors of commission. So, for both types of material, errors of commission were made more often with the auditory than with the visual stimuli. Significantly more correct detections of missing elements were made visually with the numeral series, but with the alphabet sequences there was no difference between the senses.

Alphabets and numeral sequences were chosen, it will be remembered, because they were considered to be overlearned to different degrees and so, it was thought, would provide two separate levels of task difficulty. A comparison of the error scores for the two types of material as pictured in Fig. 2 shows that this expectation was successfully fulfilled. Roughly two times as many errors were made with the alphabet sequences as with the numerals.

The results from individual *Ss* looked at separately parallel closely the averaged result shown in Fig. 2. With the numeral sequences, 7 out of 10 *Ss* made better scores visually than aurally, 1 exhibited no difference, and 2 made better scores aurally. With the alphabets, 3 *Ss* performed better visually, 5 showed no systematic bias, and 2 displayed better performance with the auditory material. The 2 who showed an aural preference for the numeral sequences were not the same 2 who showed an aural preference for the alphabet sequences.

Simultaneous trials.—In considering the simultaneous performances, the important data are not absolute error scores on the simultaneous trials but rather error scores for the simultaneous trials relative to nonsimultaneous operation. Since an increase in errors for the simultaneous detections is to be expected because of the conflicting nature of the material, a useful way of considering these data is in terms of the average increase in errors (over nonsimultaneous performance) as a function of the number of elements missing. Figure 3 shows the data from the simultaneous trials plotted in this manner. These curves were obtained by subtracting the average error scores obtained under nonsimultaneous conditions of presentations from the appropriate scores for the simultaneous trials. For example, when the auditory alphabets were presented with the visual numerals (one of the four simultaneous conditions mentioned previously), the error scores for the auditory alphabets on the nonsimultaneous trials were subtracted from the error scores obtained for the auditory alphabets. A like procedure was used for the visual numerals, and the two remainders were plotted on the same coordinates. The result is the pair of functions shown at the top of Fig. 3. Errors of omission and errors of commission are plotted separately.

In looking first at the errors of omission in Fig. 3, the most clear-cut result concerns the manner in which the type of material used affected the increase in detection errors. Both cases that combined a difficult task and an easier one displayed significant differences in increase of errors with the simultaneous material. The interesting thing is that in both instances the easier task was subject to the greatest increase in errors. Those trials that paired tasks of nearer equal difficulty showed no significant differ-

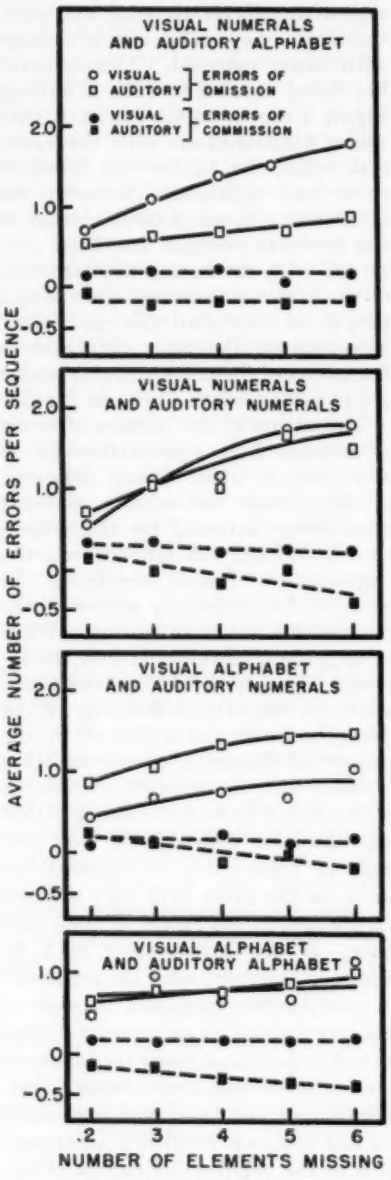

FIG. 3. Average increase in errors per sequence as a function of the number of elements missing for all simultaneous conditions

ences in error increases from nonsimultaneous to simultaneous performance.

The errors of commission presented in Fig. 3 provide an interesting analysis, likewise. The significant differences in the increase of this type of error follow the channel stimulated rather than the kind of material utilized in the detection task. The greatest increase in errors was made in every case with the visual material, which gave the fewest errors under nonsimultaneous conditions. In most instances, as can be seen from the negative scores in Fig. 3, there was a decrease in auditory errors from nonsimultaneous to simultaneous operation.

TABLE 1*

ANALYSIS OF VARIANCE OF INCREASED ERRORS OF OMISSION DUE TO SIMULTANEOUS STIMULATION

Source	*df*	Mean Square	*F*	*p*
Between materials (*M*)	1	20,478.1	103.00	>.001
Between channels (*C*)	1	23.5		
Between subjects (*S*)	9	976.5	4.90	.019
Interaction (*M*×*C*)	1	351.6	1.76	
Interaction (*M*×*S*)	9	571.1	2.87	
Interaction (*C*×*S*)	9	612.5	3.07	
Triple interaction (*M*×*C*×*S*)	9	199.2		
Total	39			

* Scores used for analysis were the average increase in errors per sequence × 100.

Two analyses of variance were performed on the foregoing results to provide a statistical analysis of the trends indicated by the graphs. In view of the fact that each *S* was used as his own control and a correlation factor was involved in the raw score used for the analysis, the method of extracting variances suggested by McNemar (6, p. 306) was utilized.

Table 1 shows the results from the analysis of variance of the errors of omission. From this analysis it is clear that the most important increase in errors due to simultaneous stimulation is caused by the type of material

TABLE 2*

ANALYSIS OF VARIANCE OF INCREASED ERRORS OF COMMISSION DUE TO SIMULTANEOUS STIMULATION

Source	*df*	Mean Square	*F*	*p*
Between materials (*M*)	1	2,433.6	3.6	
Between channels (*C*)	1	9,985.6	15.5	.006
Between subjects (*S*)	9	2,790.9	4.3	.035
Interaction (*M*×*C*)	1	62.5		
Interaction (*M*×*S*)	9	1,472.1	2.3	
Interaction (*C*×*S*)	9	993.8	1.5	
Triple interaction (*M*×*C*×*S*)	9	643.8		
Total	39			

* Scores used for analysis were the average increase in errors per sequence × 100 + 100.

rather than by the sensory channel. The increase in errors was greater when the detection task was an easy one than when it was more difficult, i.e., the efficiency in detecting deleted elements deteriorated more under conditions of simultaneous stimulation when the task was an easy one than when it was more difficult. On the other hand, Table 2, which analyzes the variances contributed by the errors of commission, shows quite an opposite result. In this case, the only significant difference, except that contributed by individual differences, derives from the channel stimulated and not from the type of material used in the test situation. The greatest increase in errors from nonsimultaneous to simultaneous stimulation is, in this case, contributed by the visual stimuli.

DISCUSSION

That there is a basic difference in absolute level of performance between vision and audition has been shown very often. Day and Beach (3), in their survey of the literature on this point, give strong support to the thesis that whichever modality dominates depends to a large extent on the type of experimental material used; and, more explicitly, previous research results suggest that vision has an advantage with difficult material, and audition may or may not have an advantage with easier material. This research has failed to confirm such a finding. Figure 2 shows, on the contrary, that vision predominated with the easier task while the harder one failed to show any differences between the modalities. Since, however, most of the previous research has been concerned with measuring the retention of the presented material over longer periods of time than this experiment encompasses, the basic difference in the nature of the tasks probably makes a direct comparison of results invalid.

The nature of the increase of errors of omission from nonsimultaneous to simultaneous observations indicates, in effect, that less missing elements were being detected by the respective modalities. In other words, the responses per channel were fewer. In view of the conflicting nature of the stimuli this is not surprising. What is surprising, however, is that the increase in errors should be determined solely by the ease or difficulty of the perceptual task and not at all by the perceptual channel involved; and, what is more, that the greatest increase in errors should be with the material that was the least difficult. It may perhaps be argued that *S*s concentrated more on the tasks that were difficult to the greater detriment of the easier ones. If the introspection of *S* be allowed, however, such an argument is considerably weakened in view of the fact that all *S*s expressed a preference for the visual tasks regardless of the type of sequence being used. Whether or not an alleged preference for one task over another is an expression of the relative difficulties of the two tasks is, of course, a debatable point. It is curious, however, that such a preference had no detectable

influence on the result of the simultaneous trials.

Some sort of end-effect might be invoked to explain the relative increase of errors with the easier material under the simultaneous conditions. The argument would run thus: Errors of omission, which contribute by far the most to the effect under discussion, are the reverse of successfully detected missing elements. If the number of successful detections was already very low for the difficult material under the nonsimultaneous conditions, then deterioration of detection scores, or increase in errors of omission, could not be relatively very great with the simultaneous tasks. Table 3 shows that such an hypothesis is extremely unlikely. Here are listed the over-all percentages of correct detections of missing elements for both levels of material difficulty and for both sensory channels. It is evident that the percentage of correctly reported letters missing from the alphabets could have dropped far below the 32% level without being critically near the zero point.

A reasonable hypothesis to advance in explanation of the large difference in errors of commission between visual and auditory trials for the nonsimultaneous experiments seems to reside in the greater referability of the visual stimuli. Although the visual sequences were presented for only a limited amount of time, *Ss* were in no way constrained from spending as much of the available time as they liked on any portion of the sequences no matter how small. Although they were requested to skim through the sequences once, rapidly, there was nothing to prevent them from looking back and checking up on elements that were missing. The fleeting nature of the auditory stimulus, on the other hand, probably caused a greater amount of uncertainty and perhaps encouraged guessing.

TABLE 3

PER CENT CORRECT DETECTIONS FOR TEN *Ss* AND FOR ALL NONSIMULTANEOUS AND SIMULTANEOUS CONDITIONS

Modality	Nonsimultaneous		Simultaneous	
	Alphabet	Numerals	Alphabet	Numerals
Visual	51.6	76.0	31.6	43.0
Auditory	54.0	71.1	31.8	39.1

This analysis can only be applied to the errors of commission made on the simultaneous trials if it is somewhat modified. Here the absolute number of errors made with the auditory material was higher than with the visual material, but relative to nonsimultaneous performance there was a large increase in visual errors and a small but consistent decrease in auditory ones. This research provides no direct answer to the problem, but it seems within the bounds of reason that the confidence with and preference for the visual task, which may be responsible for the relative absence of errors in the nonsimultaneous trials, might also lead *Ss* into a false sense of security when they are obliged to report on the simultaneously presented sequences.

SUMMARY AND CONCLUSIONS

The relative performance of vision and audition under conditions of simultaneous stimulation was investigated for rapid scanning tasks considered to exhibit two levels of difficulty. The *Ss* were required to detect elements missing from alphabets and numeral sequences when such sequences were presented simultaneously to the two senses in one of four combinations—visual alphabet with auditory alphabet, visual numeral with auditory numeral, visual alphabet with auditory numeral, or visual numeral

with auditory alphabet. Performance on the simultaneous trials was compared with nonsimultaneous performance of the same conditions. Analysis of performance was made on the basis of two types of errors. These were errors of omission, or deleted elements which were not reported as missing, and errors of commission which were elements reported as missing which were not in fact missing. The following conclusions may be drawn:

1. For nonsimultaneous operation more errors of omission were made with the alphabet sequences than with the numeral sequences. Significantly more auditory than visual errors of omission were committed with the numeral sequences. With the alphabet sequences there was no significant difference between the visual and auditory errors of omission.

2. For nonsimultaneous operation four to five times as many errors of commission were made with the auditory sequences as with the visual sequences.

3. Increase in the number of errors of omission from nonsimultaneous to simultaneous trials because of the conflicting nature of the material was not significantly greater for one modality or the other. Significant increases that did occur were attributable to the type of material used, i.e., numerals or alphabets. When a relatively easy task was paired with a relatively difficult one, significantly more errors were made with the easier task.

4. Increase in the number of errors of commission from nonsimultaneous to simultaneous trials was significantly greater for the visual trials than for the auditory.

(Received for priority publication June 30, 1952)

References

1. Carver, M. E. Listening versus reading. In H. Cantril & G. W. Allport (Eds.), *The psychology of radio*. New York: Peter Smith, 1941, 159–180.
2. Cheatham, P. G. A comparison of the visual and auditory senses as possible channels for communication. Air Materiel Command, United States Air Force, Technical Report No. 5919, 1950.
3. Day. W. F., & Beach, B. R. A survey of the research comparing the visual and auditory presentation of information. Air Materiel Command, United States Air Force, Technical Report No. 5921, 1950.
4. Harris, J. D. *Some relations between vision and audition*. Springfield: Charles C Thomas, 1950.
5. Larsen, R. P., & Feder, D. D. Common and differential factors in reading and hearing comprehension. *J. educ. Psychol.*, 1940, **31**, 241–252.
6. McNemar, Q. *Psychological statistics*. New York: John Wiley & Sons, 1949.
7. Pohlmann, A. *Experimentelle Beiträge zur Lehre vom Gedächtnis*. Berlin: Gerdes & Hodel, 1906.
8. Woodworth, R. S. *Experimental psychology*. New York: Henry Holt, 1938.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The issues listed below are available for sale

Year	Volume	Available Numbers						Price per Number	Price per Volume
1916	1	1	2	3	4	5	6	$1.25	$7.00
1917	2	1	2	–	4	5	–	$1.25	$5.00
1920	3	–	2	–	4	5	–	$1.25	$3.75
1921	4	1	–	3	4	5	6	$1.25	$6.25
1922	5	–	–	3	4	5	–	$1.25	$3.75
1923	6	–	–	–	–	–	–	——	——
1924	7	1	2	–	–	5	–	$1.25	$3.75
1925	8	1	2	3	4	–	6	$1.25	$6.25
1926	9	–	2	3	–	5	6	$1.25	$5.00
1927	10	–	–	3	4	5	6	$1.25	$5.00
1928	11	1	2	3	4	5	6	$1.25	$7.00
1929	12	1	2	3	4	5	6	$1.25	$7.00
1930	13	1	2	3	4	5	6	$1.25	$7.00
1931	14	1	2	3	4	5	6	$1.25	$7.00
1932	15	1	2	3	4	5	6	$1.25	$7.00
1933	16	1	2	3	4	5	6	$1.25	$7.00
1934	17	1	2	3	4	5	6	$1.25	$7.00
1935	18	1	2	3	4	5	6	$1.25	$7.00
1936	19	1	2	3	4	5	6	$1.25	$7.00
1937	20	1	2	3	4	5	6	$1.25	$7.00
	21	1	2	3	4	5	6	$1.25	$7.00
1938	22	1	2	3	4	5	6	$1.25	$7.00
	23	1	2	3	4	5	6	$1.25	$7.00
1939	24	1	2	3	4	5	6	$1.25	$7.00
	25	1	2	3	4	5	6	$1.25	$7.00
1940	26	1	2	3	4	5	6	$1.25	$7.00
	27	1	2	3	4	5	–	$1.25	$6.25
1941	28	1	2	3	4	5	6	$1.25	$7.00
	29	1	2	3	4	5	6	$1.25	$7.00
1942	30	1	2	3	4	5	6	$1.25	$7.00
	31	1	2	3	4	5	6	$1.25	$7.00
1943	32	1	2	3	4	5	6	$1.25	$7.00
	33	1	2	3	4	5	6	$1.25	$7.00
1944	34	1	2	3	4	5	6	$1.25	$7.00
1945	35	1	2	3	4	5	6	$1.25	$7.00
1946	36	1	2	3	4	5	6	$1.25	$7.00
1947	37	1	2	3	4	5	6	$1.25	$7.00
1948	38	1	2	3	4	5	6	$1.25	$7.00
1949	39	1	2	3	4	5	6	$1.25	$7.00
1950	40	1	2	3	4	5	6	$1.25	$7.00
1951	41	1	2	3	4	5	6	$1.25	$7.00
	42	1	2	3	4	5	6	$1.25	$7.00
1952	43	By subscription, $14.00 (2 vols.)						$1.25	——
	44								
		List price, Volumes 1 through 42							$269.00
		30% discount							80.70
		Net price, Volumes 1 through 42							$188.30

Information about the Journal of Experimental Psychology: as noted in the table, the journal was not published in 1918 and 1919; from 1937 to 1943 it was published at the rate of two volumes per year; from 1944 to 1950, in one volume per year. Beginning in 1951 there are two volumes a year.

Information about prices: when all numbers are available, the price per volume throughout the series is $7.00. For incomplete volumes, the price is $1.25 for each available number. For foreign postage, $.25 per volume should be added. The American Psychological Association gives the following discounts on orders for any one journal:

10% on orders of $ 50.00 and over
20% on orders of $100.00 and over
30% on orders of $150.00 and over

Address orders and subscriptions to

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue N.W. Washington 5, D. C.

Not Just Stimulation
On or Off
Yes or No

But controllable, measurable amounts of stimulation can be applied with our Model 226B Electronic Stimulator. You can easily use stimulation as the independent variable in experiments with animals and human beings. Write for our bulletin.

C. J. APPLEGATE & CO.

1816 Grove Street Boulder, Colorado